

Challenges of the Grid

Science policymakers worldwide are waking up to the opportunities of 'the Grid' — a supercomputing network transforming many disciplines. But it poses new organizational challenges for researchers and their institutions.

After the ILOVEYOU virus outbreak, a cartoon in the newspaper *Le Monde* portrayed a company boss thanking his assembled staff profusely for being so computer illiterate as to be still ignorant of how to open an e-mail attachment. Thankfully, the computing *savoir faire* of the average scientist is considerably greater. Yet there is an irony. The desire of scientists to collaborate inspired the invention of the World-Wide Web at CERN. But whereas the Internet has since infiltrated every inch of the planet, its use for widespread scientific collaboration — apart from a web-wise minority — remains largely rudimentary compared with the potential offered.

For most scientists, the web remains a place to search for information. For many, building a laboratory home page is the closest they come to interacting with the medium, and most of these sites are static; the electronic medium may mean that they are immediately accessible to all, but they provide users with little interactivity. Highly interactive laboratory pages, where, for example, any member of a community can edit databases and pages online, remain the exception.

One reason can be found in the dominance of proprietary software. Microsoft's 2000 release of its Office suite, for example, brings collaborative web working tools to such workhorses as Word and Excel. But in order to benefit, every user must have Office 2000 and Microsoft server extensions. Barring a considerable investment in new software, this makes a collaborative web environment difficult even in an intranet where everyone is using Microsoft products and PCs. In the Internet environment, where users may be using Macs, Linux or Unix operating systems, and a multitude of programs, such closed systems are not a solution.

Open and free

As it happens, simple open-source and highly interoperable tools for collective working exist, and are often superior to their commercial counterparts. Apart from the costs of hardware, it is possible to build an excellent site, for example, using the Linux operating system, an Apache WWW server and the MySQL database. To this one can add such features as WikiWeb, a collaborative website that can be edited by anyone (<http://www.ensembl.org/Docs/wiki/html/EnsemblDocs/GettingStartedWithWiki.html>), a simple web database interface (<http://www.ofifc.org/Eli/ASP/GenericArticle.asp>) or a means for users to upload any sort of file online (<http://www.dougdean.com/EZsiteUploadSite.htm>) — all are free, with most requiring only a minimal knowledge of computing and programming languages.

Plans that will encourage a more sophisticated and highly productive use of networks are now afoot. The National Computational Science Alliance in the United States and, more belatedly, the European Union, are enthusiastically embracing an ambitious architecture for the next-generation Internet: a 'computational Grid'. The concept is elegant. Few of us have the benefit of a supercomputer next door. In the Grid, any scientist could plug their computer into a high-speed Internet and immediately share huge supercomputing resources — and in particular supercomputer 'clusters' of cheap PCs — distributed across a high-speed Internet. The increasing occurrence of tera- and petabyte/flop datasets, data flows and computations in science is

creating a real demand for such computing power in many disciplines.

But perhaps the most significant vision is one of an Internet where all sorts of datasets, visualization tools and software can communicate in the same language. This could be achieved through a software layer — with the uninspiring name of 'middleware' — that would allow anyone on their desktop to draw on datasets wherever they were held on the Internet without having to worry about formats or protocols. The Internet becomes the computer.

Collaboratories

At a recent 'E-science' workshop hosted by the European Commission, the backing for grids from disciplines as diverse as genomics, particle physics, space science and meteorology was, to say the least, wildly enthusiastic. The vision is well-founded. In short, it is one where people, computers and services no longer operate in isolation, but are shared and made available to the entire community; individual laboratories would give way to international 'collaboratories' where geographically dispersed groups would work seamlessly on shared datasets; it is already happening (see *Nature* **402**, C67–C70; 2000).

Plans for grids are dominated by the high-performance supercomputing community, mainly physicists. Their leadership promises benefits to other less computing-savvy areas of science. The immediate next steps in the Grid projects are eminently laudable, for example to build high-speed networks, impose order on disorder by allowing databases and other applications to speak the same language.

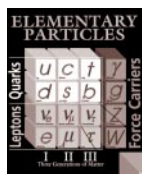
But there is a danger in such top-down approaches, namely that the technology risks becoming an end in itself, removed from the real needs of diverse grassroots users. In the early days of physics, and more recently in genomics, the major strides — such as Ensembl (see page 333) — have largely been driven by a 'hacker culture' where computer-literate scientists designed and built information-technology systems to satisfy urgent and pressing bottom-up research needs. The challenge will be to merge these two cultures.

Using current browsers, anyone can access dumb HTML documents. Grid browsers will allow scientists just as easily to access computing resources, databases and tools. Using XML, CORBA and other languages and protocols, it is already possible for websites and online databases to communicate, interpret and understand one another's content. It is up to individual communities to organize themselves to take advantage of this opportunity.

The Grid concept is without doubt the way to go, and will benefit scientists greatly. But to maximize the benefits of the next-generation Internet in an optimal, sustainable and rapid way, it cannot be left to either the enlightened vision of a few computing gurus or the heroic effort of entrepreneurial scientists at the bench concerned with real research problems, doubling up as computer scientists.

It is a common, and perhaps unjust, maxim that whenever you want to do something progressive in computing, the last place to look is your institution's IT department. Nevertheless, one way or another, all institutions involved in science need to give priority to how modern information technology can best serve them, and equip themselves with the necessary budgets and competence. ■





A small matter
Physicists find the
'final' fundamental
particle
p334



In tune
Clinton and Blair say
GM regulation must
be 'science based'
p335



Ethical research
New code proposed
for collaborative
research projects
p337



Head start
Neurobionic
institute unveiled
in Hanover
p339

Emissions targets 'unrealistic' says US climate change body

Colin Macilwain, Washington

US climate-change activists are up in arms again. But for once it is not against the fossil-fuel lobby. This time the focus of their ire is one of their strongest supporters — the Pew Center on Global Climate Change.

One of the main organizations working for cuts in US greenhouse-gas emissions, the Pew Center has issued a statement saying that the emissions targets set by the 1997 Kyoto Protocol are unrealistic and will eventually have to be renegotiated. Eileen Claussen, the centre's president and former chief negotiator on climate change issues at the US state department, says that "few national governments" will reach their targets for 2008–12.

"There is nothing wrong with ambitious targets, but they have to be grounded in reality," Claussen said in remarks delivered privately in London last month, and repeated publicly in Washington last week at a meeting of scientists and economists who advise the Pew Center.

"The targets in the Kyoto Protocol cannot and will not be met on the established timetable in the United States and elsewhere," said Claussen. "By adhering to unrealistic targets that will be very difficult, if not impossible to meet, we provide the Protocol's opponents with additional ammunition in their efforts to shoot the treaty down."

The Pew Center was established in 1998 by the Pew Charitable Trusts to educate the public on climate-change issues and to encourage reductions in greenhouse-gas emissions. It operates in partnership with a group called the Business Environmental Leadership Council, which includes Boeing and DuPont, but receives no financial support from it.

Claussen still believes that "the good qualities of the Kyoto Protocol vastly outweigh its flaws" and she wants negotiators, who will meet in The Hague this November to finalize the protocol, to implement firm rules on how the treaty will operate before considering revisions to the targets. "If, after ensuring that the framework reflects these priorities, we need to renegotiate the targets or timetables — and I suspect we will — then so be it."



High and dry: avoiding droughts like this one in Florida could require achievable emissions targets.

The remarks were welcomed by some climatologists at the Washington meeting. "I agree with everything she said," says Tom Wigley, senior scientist at the National Center for Atmospheric Research at Boulder, Colorado. "The targets are too much, too soon. She is echoing what many economists believe, that a more moderate approach is

going to be a much better approach."

But Claussen's blunt characterization of the targets has alarmed environmental groups, who believe that the Kyoto Protocol is the main driving force for global action by governments and industry to cut emissions.

Claussen's pessimism is not shared, at least in public, by other groups who support action

Ensembl gets a Wellcome boost

Declan Butler

Britain's Wellcome Trust announced last week that it is contributing £8.8 million (\$13.4 million) to Ensembl, the decentralized annotation project for the human genome.

The funding, over five years, is a vote of confidence for Ensembl's vision of sequence annotation being distributed across multiple computer servers worldwide, rather than in a centralized server such as GenBank.

A joint venture between the Sanger Centre and the European Bioinformatics Institute, both at Hinxton near Cambridge, Ensembl aims to provide a single reference map of the human genome on to which participating scientists can anchor 'annotations' describing the predicted function and structure of gene sequences.

This Distributed Sequence Annotation System (DAS), which is being developed by

Lincoln Stein, a bioinformatician at Cold Spring Harbor Laboratory in New York, will be rolled out next month at the International Conference on Intelligent Systems for Molecular Biology in La Jolla, California.

A software 'plug-in' will then be made available to allow users to zoom in on any part of the genome. As well as accessing details on the DNA sequence, together with author information, they will be able to compare apparent contradictions between different third-party annotations. The results will be returned in standard formats from all participating DAS servers, so that the graphics and displays look the same.

The first two DAS servers to join the scheme will be at Ensembl and the Oak Ridge National Laboratory in Tennessee. ■

♦ <http://www.ensembl.org/>

♦ <http://stein.cshl.org/das/>



Claussen: targets for emissions must be realistic.

to cut carbon emissions. "We're making a lot of progress here," says Dan Lashof, head of the climate-change programme at the Natural Resources Defense Council, pointing to recent bipartisan calls in the Congress for reductions in power-plant emissions.

Jennifer Morgan, director of the World Wildlife Fund's climate-change campaign, also believes the Kyoto Protocol can work. "We should be focusing on a domestic plan to meet the targets, not on changing them," she says.

But the Global Climate Coalition, the industry group that has opposed binding action to cut carbon emissions, said that Claussen's statements bore out its early warnings about the Kyoto Protocol. "We agree with her that the targets are simply unrealistic," says Connie Holmes, chair of the coalition's board. "In the last three years, they have become more unrealistic — and not just for the United States."

► http://www.pewclimate.org/media/transcript_riia.html

Proposal for US patent office could help cut waiting time

Paul Smaglik, Washington

Anyone who has recently applied for a US patent — particularly in high-tech fields or biotechnology — will have waited a long time to learn of their request's fate. A congressman is now proposing to address this problem by allowing the United States Patent and Trademark Office (USPTO) to keep all the fees that it raises.

The move would help to increase the number of examiners — without this, warns the office, the backlog of applications is likely to grow even longer. Applications have increased by 12% in the past year alone, and the time for processing each one has grown from 18 months in 1995 to 26 months now.

Up to now, the patent office has always handed over a portion of its income to the government. But the precise percentage has varied, creating difficulties with long-term hiring commitments.

In the next financial year, which begins in October, the USPTO expects to take in about \$1.2 billion. But the House, Senate and President Bill Clinton's administration disagree

over how much the agency should keep. The House is suggesting \$905 million, whereas the Senate last week voted to allow it to retain all but \$33 million; the president's budget calls for the USPTO to give up \$113 million.

Representative Howard Coble (Republican, North Carolina) has put forward a bill that would allow the USPTO to spend everything it earns. The administration has not commented on this proposal. But in a letter to Coble earlier this summer, USPTO commissioner Todd Dickinson warned of dire consequences if the House's budget were adopted, as the office would have to impose a hiring freeze and reduce overtime.

Demand for patents is especially heavy in genomics, where there is already a backlog of 20,000 applications for patents on genes, says Lila Feisee, a spokesperson for the Biotechnology Industry Organization.

Analysing these applications requires more people and better software, adds Feisee, a former USPTO examiner. "There's such an outcry over gene patents now that it's important to do it right," she says.

Physicists celebrate detection of elusive 'final' particle

Colin Macilwain, Washington

High-energy physicists are declaring 'full house' after detecting the last remaining basic building block of matter predicted by the Standard Model. A team at Fermilab outside Chicago has seen the elusive tau neutrino for the first time.

The DONUT (Direct Observation of the Nu Tau) team of 54 physicists from the United States, Japan, South Korea and Greece announced last week that they had found the kinked, millimetre-long decay tracks of four tau leptons — produced when an iron nucleus and a tau neutrino collide — after analysing millions of such possible events.

The Standard Model of fundamental particles and their interactions predicts the existence of 12 'fermions' — the basic building blocks of matter — and four 'vector bosons', which carry force. The tau neutrino, which belongs to a class of fermions called leptons, was the last to evade detection.

"Finding the tau neutrino is very important and exciting," says Martin Perl of Stanford University, California, who won the 1995 Nobel Prize in Physics for his discovery of the tau lepton. "DONUT was not an easy experiment, and now it opens a whole new world. We might have a chance of learning more about all other particles."



DONUT delight: spokesman Lundberg describes discovery as 'the proverbial needle in a haystack'.

In particular, the finding may assist the burgeoning field of neutrino physics. Early results from the SuperKamiokande experiment in Japan suggest that neutrinos may have mass (see *Nature* 391, 123; 1998). To test this, physicists in Japan, Europe and the United States are planning experiments to observe the oscillation of one type of neutrino to another.

The DONUT experiment began in 1997, by sending a high-energy proton beam from Fermilab's main particle accelerator, the Tevatron, into a block of tungsten.

Researchers filtered a beam of neutrinos out of the resulting mêlée, and this was passed through a target of iron sandwiched with layers of emulsion.

The DONUT team identified 1,000 events of interest and took the emulsion plates, which work like three-dimensional photographic film, to Nagoya University in Japan for analysis. "It was the proverbial needle in a haystack," says Fermilab's Byron Lundberg, spokesman for the experiment.

High-energy physicists are now switching their attention to the Higgs mechanism, an extension to the Standard Model, which explains how particles acquire mass. This involves exchanges of a particle called the Higgs boson, the search for which will resume at a revamped Tevatron next year. CERN, the European Laboratory for Particle Physics near Geneva, will join the hunt in 2005 with its Large Hadron Collider.

Meanwhile, some researchers are excited by 'supersymmetry', a theory outside the Standard Model. This predicts that every fermion has a supersymmetric boson counterpart, and vice versa. Experiments at CERN's Large Electron-Positron collider have hinted at a counterpart to the bottom quark — although previous hints have turned out to be statistical fluctuations.

Summit leaders fail to bridge GM food split

David Dickson

The G8 group of the leaders of the world's largest industrialized nations agreed at their annual meeting last weekend on the need to accelerate the application of a "science-based, rule-based" approach to ensuring food safety — including the regulation of genetically modified (GM) crops.

But the meeting failed to bridge a split over political strategy on handling GM foods. US President Bill Clinton, backed by British Prime Minister Tony Blair, argued that their future regulation "should be based on science". In response, Blair was reported to have said that consumers needed "the best science available. You get the real facts, not the prejudices."

In contrast, other members of the European Union called for a more 'precautionary' approach, and a greater involvement of consumer associations and other interest groups in decision-making on the safety of GM foods — neither of which appeal to the US biotechnology industry.

The final communiqué of the three-day meeting, held in the Japanese city of Okinawa, attempted to reach a compromise. The document pledges the support of the G8 leaders to efforts "to achieve greater consensus on how precaution should be applied to food safety in circumstances where available scientific information is incomplete or contradictory".

But, in a further indicator of the split, the meeting failed to endorse a proposal to set up an international forum to assess both the science and the social implications of GM foods. This had been suggested by a number of senior scientists, including John Krebs, head of Britain's Food Standards Agency, at a meeting in Edinburgh earlier this year, organized in response to concerns about GM crops voiced at last year's G8 summit (see *Nature* 404, 112; 2000).

Supporters of the proposed forum had argued that such a body could take its model from the successful operation of the Inter-governmental Panel on Climate Change, using the scientific dimensions of the debate as a framework within which to construct a dialogue between participants with widely differing views.

But the United States, keen to limit opposition to its agribiotech industry, is said to have remained firmly opposed to any such initiative. Reflecting this view, a proposal that such a committee be set up by the Paris-based Organization for Economic Cooperation and Development was rejected earlier this year.

Similarly, the proposal was merely "noted" in the summit's final communiqué, with the G8 leaders only promising to explore "in consultation with international

organizations and interested bodies including scientific academies" how to integrate "the best scientific knowledge available" into a global consensus on biotechnology.

In the same document, the G8 leaders also welcomed the "nearly complete" mapping of the human genome. They called for "further rapid release of all raw fundamental data on human DNA sequences as such" — thus aligning themselves with the bilateral appeal on the same issue made by Clinton and Blair in March (see *Nature* 404, 324–325; 2000).

In addition, the leaders emphasized "the importance of pursuing the post-genome-sequence research on the basis of multilateral collaboration". And although declining to endorse proposals that are said to have been put to the meeting for a new initiative to harmonize international rules on gene



Odd two out? Clinton and Blair oppose Europe's more 'precautionary' approach to biotechnology.

patenting, they did agree on "the need for a balanced and equitable intellectual property protection for gene-based inventions". ■

UK science plans spending spree

Natasha Loder, London

Information technology, bioinformatics, nanotechnology and post-genomic research are all set for a funding boost from the UK government. They are some of the areas of interest identified by the UK research councils in their bid for a share of the extra spending on science announced last week (*Nature* 406, 225; 2000).

Research council priorities in bids for new money

Cross-council

- Bioinformatics
- Post-genome research
- Advanced computing and 'the Grid'

Particle Physics and Astronomy Research Council

- Application of the Grid to Large Hadron Collider
- UK membership of European Southern Observatory*

Engineering and Physical Sciences Research Council

- Nanotechnology and quantum computing
- Photonics

*This requires the government to fund a separate request for the one-off entry fee

Biotechnology and Biological Sciences Research Council

- Functional and structural genomics
- Application of the Grid to accessing and manipulating data generated through genomics
- Nanobiotechnology and bioscience engineering

Medical Research Council

- Application of genomics to human health
- Health of the public

Natural Environment Research Council

- Environmental genomics
- Sustainable land use
- Application of the Grid to climate modelling

Over the next three months, the research councils will battle it out to see how much of these extra funds each will receive. The precise allocation will not be announced until the autumn. Overall, the science budget will grow by more than 4% over the previous year in each of the next three years.

In addition to next year's increase, the government is to provide an extra £50 million (US\$76 million) fund for recruiting and retaining top academics. Although little has yet been announced about how this fund will operate, it might be used to top-up the salaries of high-performing academics in a bid to make their income more internationally competitive, reducing the risks of them leaving the country.

The Association of University Teachers trade union says that it expects payments from the fund to be linked to improvements in management performance. A 23% increase to the stipends of research students has already been announced (*Nature* 406, 113–114; 2000).

The overall settlement has been warmly greeted both by the individual research councils and by the Royal Society. "The broad message is we are very happy," says Bob Ward, manager of the science advice section of the Royal Society. "We were amazed that the government appears to have addressed all the main areas we were worried about."

But universities must still find £1 for every £3 of taxpayers' money invested in research infrastructure — an attempt to encourage them to strengthen their links with business. ■

Legal protests prompt DNA primer release

Paul Smaglik, Washington

A leading supplier of DNA testing kits widely used in legal actions has promised to make public its closely held proprietary information on the sequence of the primer used in the kits.

In a number of recent cases, defence attorneys have successfully challenged the validity of DNA testing on the grounds of lack of access to such information. The companies that make the tests have withheld the information through commercial confidentiality. However, the market for such tests could diminish if they are rendered legally invalid.

On 18 April, for example, a state court in Vermont ruled that DNA results were inadmissible as evidence in a criminal case because the two main companies that had provided the tests — Promega and PE Biosystems — had not made their primer sequences available. Last week, Promega said it will divulge those sequences in future.

To establish whether crime-scene DNA samples match those from suspects, forensic scientists analyse DNA from points in the genome at which the letters in the genetic code repeat, focusing on 13 sites where the number of repeats varies highly between individuals. The primers are designed to



Mozer: revealing sequences will remove doubt.

bind to the sequences flanking these repeating regions, allowing scientists to multiply the repeated regions over and over again using the polymerase chain reaction.

This amplification process enables scientists to count the number of repeats in both the evidence and the suspect's DNA sample. If the number of repeats in each of the 13 regions matches, it is highly likely that the two samples came from the same individual.

Some defence attorneys have recently argued successfully that, in principle, the primers could contaminate the DNA being

amplified, invalidating the test. Without knowing the sequence of the primers, however, courts cannot determine whether the primer sequence is present in the samples.

Promega researcher Tom Mozer says the chances of contamination are extremely remote. But defence attorneys have used the lack of disclosure to introduce doubt into the proceedings. Releasing the primer sequences will help dispel that doubt, Mozer says.

Arthur Eisenberg, director of the DNA Identity Laboratory at the University of North Texas Health Science Center in Fort Worth, agrees that contamination from primers is rare. "It's an argument that the defence has used to try to keep this testimony out of court," he claims.

While applauding Promega's decision to release the information, he speculates that the company may be hoping forensic scientists will turn to its tests rather than its competitors' because they stand a better chance of being accepted in court.

And although Promega is divulging valuable intellectual property, Eisenberg doubts that many scientists will bother to copy them all. Although it is feasible to create one or two primers by hand, a complete pre-prepared set is far more useful. ■

ANDY MANIS

Review panel assails Brussels research bureaucracy

Quirin Schiermeier, Munich

European scientists have long griped about the bureaucracy and delays involved in extracting funds from the European Commission. Their complaints have now been endorsed by an independent advisory panel that says improvements are essential if the commission is to achieve its goals.

A survey carried out for the panel found that only 70% of participants in the commission's Framework research programmes thought the benefits had

outweighed the costs. Two-thirds complained that the application process was too slow and/or costly, and 45% found the application procedures difficult to follow.

The panel says that the Framework programme is unable to meet the coordinated science-policy goals outlined by the heads of government at a summit meeting in Lisbon last spring, intended to make Europe the world's most dynamic and competitive knowledge-based economy.

The commission sets up the panel every five years to evaluate its Framework programmes. Academics and industrialists from 11 countries carried out the 1995–99 assessment, chaired by Joan Majó, a former Spanish minister for industry.

The panel's main recommendation is that the administration of the Framework programme should be aligned with the practical needs of European scientists and industry. Its criticism of application procedures is based on the analysis of 2,275 responses to a survey of applicants to the third and fourth Framework programmes.

Scientists participating in the current fifth Framework programme (FP5) have similar complaints. In some sub-programmes, less than 20% in some sub-programmes, many

researchers wonder whether applying is worth the effort (see *Nature* 404, 695; 2000).

Panel member Yves Farge, special adviser to France's Centre National de la Recherche Scientifique, says the programmes' inefficiency is due mostly to cumbersome administrative procedures. "It is time now to react and to simplify the rules," he says.

Commission officials are already exploring improvements to the application and evaluation procedure. Some of these will be tested next February in a pilot call for proposals in bioethics and the socio-economic aspects of biological research.

The commission is also considering changing the rule that FP5 grant proposals must be anonymous during the first stage of evaluation. Although intended to prevent biased evaluation, many scientists say that it has serious drawbacks. For example, the inability to cite their previous publications makes it hard for research groups to show that their work is truly innovative.

The panel also calls on the commission to encourage more proposals for high-risk, high-return projects. "The stock of groundbreaking new ideas developed in the 1990s will not last for ever," says Farge. ■

♦ http://www.cordis.lu/fp5/5yr_reports.htm



Call for North/South code of research ethics

Declan Butler, Paris

The balance of power in research collaboration between scientists in industrialized and developing countries could be redressed by adopting an international code of ethics, says the head of a French research agency.

Philippe Lazar, president of the Institut de Recherche pour le Développement (IRD), the French national agency for scientific research for development, has drawn up a draft text setting out a series of broad principles (see below), and is talking to the French government about getting it adopted by all French research organizations.

Lazar has informally presented it to several other countries worldwide. Ultimately, an international declaration could be one possibility, he says, though at present the proposals are essentially a list of good intentions and need more work on the details. The declaration would not be legally binding — but nor is the Universal Declaration of Human Rights, Lazar points out, and adoption of a text would focus minds on the issues.

The plan has been welcomed by the scientists, legal and ethical experts, and donor agencies contacted by *Nature*. But their views differ widely on what its scope should be, and the feasibility of translating it into practice.

A plan to foster scientific cooperation between the industrialized North and the developing South comes at just the right time, says Noëlle Lenoir, a member of France's constitutional court and president of the European Group on Ethics in Science and New Technologies. Genomics, population genetics and medical genetics must be global, she says, hoping that a declaration could narrow the widening technology and research gap.

"I often encounter suspicion among researchers in the South towards pushy potential collaborators from the North —



In the clear: the code would protect legitimate specimen collecting, such as this Dutch/Brunei project.

suspicions often held with good reason," says Chris Dye, from the World Health Organization. "If fears of specimen and data poaching, and so on, can be allayed by a new code of professional conduct, so much the better."

Pierre Druilhe, a malaria researcher at the Institut Pasteur in Paris who is usually sceptical of such political initiatives, commends Lazar for taking the initiative. "The concept is good: it's about striking a new deal with developing countries," says Druilhe, who spends much of his time in these parts of the world.

But putting it into practice will be much more difficult, he says: enforcing equal conditions of health and safety at work, for example, would be almost impossible. "The text is not perfect, but if you don't make a start we will never get there."

Others question whether a text singling out North/South issues is the best approach. The text should apply to any research collaboration,

say both Gurinder Shahi, executive director of the Asia-Pacific International Molecular Biology Network, and Lincoln Chen, executive vice-president of the Rockefeller Foundation.

"Scientific, financial and other imbalances in particular associated with rich versus poor countries and their scientists and institutions could be addressed within this overall framework," adds Chen. But he agrees that adoption of guidelines would heighten sensitivity to ethical issues.

Richard Lane, head of international programmes at the UK Wellcome Trust, says the guidelines are novel in their broad scope, but they pay insufficient attention to broader questions such as who benefits from such research, or the fact that local research needs are often very different from those of Northern partners. The guidelines also assume that developing countries have sufficient expertise to negotiate effectively, says Lane, whereas in reality they suffer from an acute lack of research administrators.

"I see the exercise as well-intentioned," says Amir Attaran, from the Center for International Development at Harvard University. His concern is that the present text focuses excessively on the 'how' of research and ignores the bigger question of 'why' — what needs the research is designed to meet.

Attaran says he is disappointed that the text does not address the need for much greater research on the science priorities of the South, such as malaria and countless other neglected tropical diseases.

Lenoir, however, sees a duty in the South too. "Scientific progress in the developing countries is not only a question of cooperation between the North and the South," she says. "It is also an education and ethics issue for developing countries themselves. This should be stressed in any guidelines."

http://www.ird.fr/fr/inst/debat/

Main points of the draft declaration

- No research in a developing country can be done without the participation of teams from that country.
- The scientific quality of research conducted in cooperation with developing countries must be the same as that in industrialized countries.
- Northern partners must help train scientific and technical staff in developing countries, and limit the risk of a 'brain drain'.
- Industrialized countries' rules relating to the planning and management of research programmes must be strictly respected in any joint projects.
- Each project must undergo systematic ethical examination, taking into account the developing country's culture.
- Health and safety conditions must be the same for everyone involved.
- All participants must be informed about every part

of the programme, particularly of any risks and any possible economic or social implications.

- Everyone involved must have right of access to the various methods of publication and be given the chance to maximize the benefit of their results.
- Partners must make a systematic effort to maximize the benefits for the populations and countries involved, without necessarily waiting for official completion.
- No cooperation-based research should be undertaken which, in the present state of knowledge, could be considered potentially harmful to populations, individuals or their environment.
- Intellectual property rights on data and results obtained must be shared fairly among the participants, in accordance with their overall contributions.

NIH urged to tighten up bookkeeping on external grants

Washington As it continues the rapid expansion of its grant programme, the US National Institutes of Health (NIH) needs to tighten up its accounting for the way that extramural grant money is spent, says a report from the powerful General Accounting Office (GAO).

The GAO finds that the biomedical research agency's annual check of the scientific and budgetary progress of each of its grantees has sometimes been inadequately documented. Five out of six NIH institutes studied by the audit failed to keep proper reports on completed grants, it finds, and the agency is not doing enough to recover unspent money from grantees.

But the GAO, which was asked by Congress to investigate the NIH's extramural programme, is mild in its criticisms of NIH accounting, and has no major complaints about how the agency is allocating its recent large budget increases.

US universities 'subsidize government research'

Washington US universities are subsidizing government-funded research by between \$700 million and \$1.5 billion a year because the money they receive for overheads falls short of their true cost, says an independent report commissioned by the government.

The report, by economists at the Rand Corporation, estimates that grants from agencies such as the National Institutes of Health and the National Science Foundation cover only 70–90% of the cost of university facilities and administration.

Regions compete for French synchrotron

Paris Ten regions in France are among the contenders vying to host the planned national synchrotron Soleil. France's minister of research, Roger-Gérard Schwartzberg, told a press conference last week that he would announce a decision by the end of August.

The leading bids include a joint submission from the Ile-de-France and Essonne, which have offered FF1 billion (\$143 million) towards the construction of the 2.5–2.7 GeV machine at Orsay, outside Paris. Aquitaine and Midi-Pyrénées have joined with four Spanish regions — Euskadi, Navarre, Aragon and Catalonia — to offer the same level of financing. The Nord-Pas de Calais, which would like to see the synchrotron built near Lille in northern France, with the aid of public and private funding from Belgium, has offered FF250 million.



Brainchild: a place where neuroscientists feel at home

Munich Perhaps hoping to be inspired by its architectural design, the new Hanover International Neuroscience Institute will occupy the world's first building to be shaped like a brain (above). The institute, funded by an international foundation called Neurobionic, was formally opened last week at EXPO 2000 in Hanover.

The institute will have 200 staff, 70 of whom will be scientists, and will focus on clinical research, including the development of neuroprostheses. Neurobionic was set up four years ago by a consortium of banks, companies and hospitals.

Jupiter's tiny 'moon' waits to be named

Washington Astronomers believe they have discovered another moon of Jupiter, the first to be found in a quarter of a century. If the finding is confirmed, astronomers at the University of Arizona and the Smithsonian Astrophysical Observatory in Massachusetts will have discovered Jupiter's seventeenth moon — and the smallest known satellite of any of the nine major planets in our Solar System.

The tiny moon — just three miles in diameter — is known only as S/1999 J1. It was first spotted in October last year by the 79-year-old, 36-inch telescope at the Kitt Peak Observatory in Arizona. The object was initially thought to be an asteroid, but subsequent observations and calculations have revealed that it could be orbiting Jupiter.

Surprise at UK plan to sell off part of Defence Ministry

London The British government has announced this week that it is to partly privatize the Defence Evaluation Research Agency (DERA), currently part of the Ministry of Defence. The news comes as a surprise following a damning report about the privatization plans from a cross-party group of Members of Parliament and unions earlier this year (*Nature* 402, 223; 2000).

The US Department of Defense had also

expressed security concerns over the proposals. DERA has an annual turnover of about one billion pounds, and three-quarters of its 12,000 staff could be moved into the private sector. The organization is to be separated by the end of the year after a period of 'shadow operation'.

Western world lags in 'difficult' physics

London Teams from Asia and the former Soviet bloc dominated last week's 31st annual Physics Olympiad. Sixty-three teams of five 16- and 17-year-olds from around the world competed in theoretical and experimental tests. China won, followed by Russia and Hungary, with India, Iran and Taiwan sharing fourth place.

Britain, which hosted the Olympiad at the University of Leicester, came fifteenth. Ken Pounds, the university's head of physics, said the results reflected a move away from teaching physics in many western schools, "partly because it is perceived to be difficult".

Spanish academics unite to tackle traditional problems

Madrid Tenured staff in Spanish universities have set up an organization known as the Platform for the Improvement of Higher Education Quality, to raise the standard of the country's research. They plan to fight for reforms to the selection process for academic staff, and to launch a debate on changes needed in the university system.

Under current university laws, contract researchers in a department are frequently given a tenured position, even if they are not the most scientifically qualified applicants. More than 90% of tenured academic positions go to internal candidates. Julio Iglesias de Ussel, State Secretary of Universities, admits that such practices are a "traditional problem of Spanish society".

Vast cost of nuclear stockpile could be cut

Washington The US Department of Energy's \$4.9 billion-a-year stewardship programme to maintain the country's nuclear weapons is one of the most expensive and least effective of the options available for managing the stockpile, says a comparative analysis of five such options published last week.

Robert Civiak, a physicist who worked as budget examiner of the programme at the White House Office of Management and Budget until last year, did the comparison for the environmental group Tri-Valley CAREs. His analysis estimates that a programme to remanufacture nuclear weapons when they reach a certain age would cost \$3.9 billion a year. One that allowed some to be withdrawn would cost as little as \$1.7 billion.

Manhattan versus Reykjavik

Where is it best to hunt for genes that underlie cancer and heart disease? Isolated populations such as Iceland's, or ethnic melting pots like the United States? And what are the technological challenges, asks Alison Abbott.

According to some, Iceland is selling its soul to the devil, and the devil is a company called deCODE Genetics. Based in Reykjavik, deCODE's product is genetic information linked, anonymously, to medical records for the country's 270,000 inhabitants — or, at least, for the majority who have not asked to be excluded from the company's database.

The relative genetic homogeneity of Iceland's population, it was thought, should make it a good place to investigate the genetic factors involved in conditions such as heart disease. But the Icelandic parliament's decision to grant deCODE privileged access to the necessary data has angered some citizens, who object to their country's gene stock being used to profit a single company.

Among geneticists, however, this ethical controversy is taking second place to a more fundamental scientific debate. Do isolated, genetically homogeneous populations provide any real advantage in untangling the multiple factors, genetic and environmental, that contribute to common killers such as cancer, diabetes, heart disease and stroke?

There is a growing feeling that they might not. But if larger studies in genetically mixed populations, such as those of Western Europe and the United States, are the way forward, some difficult questions must be answered. What methods of analysis should be used? And can existing technologies cope with the huge amounts of genotyping that may be required? "We need a more elegant solution than just forcing large numbers through," says Eric Lander, whose genomics centre at the Massachusetts Institute of Technology's Whitehead Institute for Biomedical Research is one of the leaders in the field.

Isolated populations provide a relatively simple genetic background against



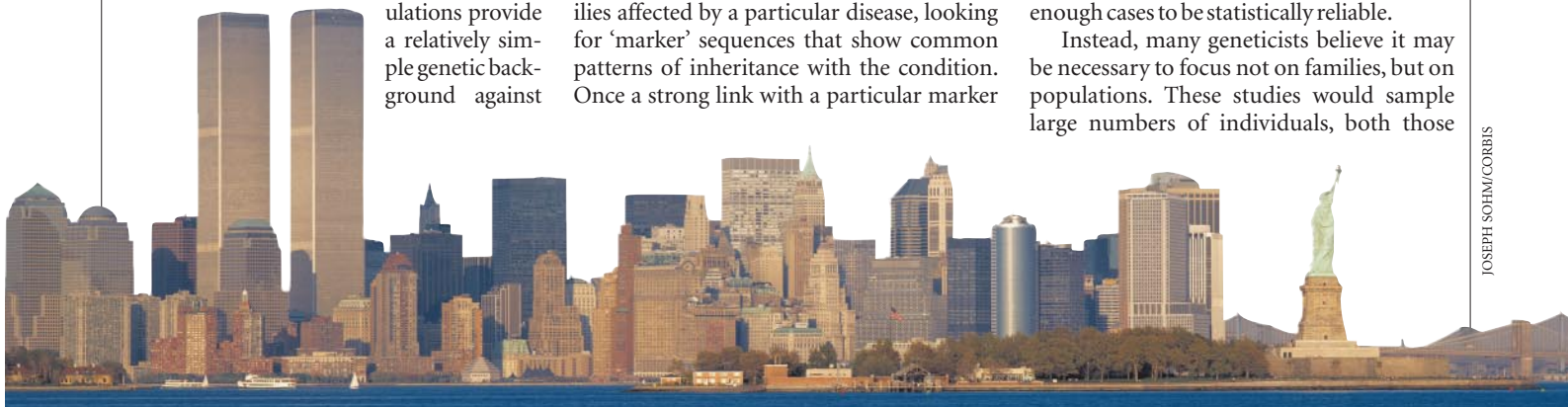
Spectrum of diversity: Andres Metspalu (right) contrasts the extreme genetic heterogeneity of Manhattan's population (below) with the homogeneity of isolated Reykjavik (above).

which to investigate the genetics of a disease. They have already proved their worth in studies of conditions caused by single defective genes. Work in Finland, for instance, led to the discovery of the genes underlying a large number of rare hereditary conditions, including specific forms of dwarfism, epilepsy and eye disorders¹. Such conditions tend to be perpetuated in the restricted gene pool of an isolated population.

Many of these discoveries have relied on a technique called linkage analysis, in which researchers study the genetic make up of families affected by a particular disease, looking for 'marker' sequences that show common patterns of inheritance with the condition. Once a strong link with a particular marker

has been found, the geneticists focus on the surrounding chromosomal region to search for the disease gene itself. But for common diseases, which might involve many genes, each with a relatively small role to play in the overall condition, linkage analysis may not be so powerful^{2,3}. Where individual disease genes exert only small effects, people carrying these genes do not always develop the disease, so studies within families might not generate enough cases to be statistically reliable.

Instead, many geneticists believe it may be necessary to focus not on families, but on populations. These studies would sample large numbers of individuals, both those



MICHAEL NICHOLSON/CORBIS

JOSEPH SOHM/CORBIS

with the disease and healthy controls. They would either look for an association between the disease and particular genetic markers, or would test hypotheses about mutations in 'candidate' genes whose normal function suggests they might be involved in disease.

Mix and match

Isolated populations such as those in Iceland and Finland were again thought to offer advantages for these 'association studies'. Many geneticists assumed that such populations would show relatively high levels of linkage disequilibrium (LD). This is the tendency for variants, or alleles, of two genetic sequences — such as a genetic marker and a nearby disease gene — to occur together within individuals more often than would be expected by chance. Researchers reasoned that in a population such as Iceland's, which expanded from a relatively small number of founders and has not experienced significant immigration, there should have been fewer opportunities for particular markers and disease genes to have become separated down the generations.

Recently, this assumption has been challenged. Last year, for instance, Leonid Kruglyak of the Fred Hutchinson Cancer Research Center in Seattle produced a theoretical model of populations with different levels of heterogeneity, and concluded that isolated populations are unlikely to be very different from more mixed populations in terms of LD⁴.

Two papers published this month in *Nature Genetics*^{5,6} provide data to support Kruglyak's conclusions. Researchers led by John Todd of the Wellcome Trust Centre for Molecular Mechanisms in Disease in Cam-

bridge studied a region of chromosome 18 in Finnish, Sardinian, British and American populations. They measured LD for genetic markers known as microsatellites and found no significant differences between the various groups. The researchers suggested that "genetic isolates like Finland and Sardinia will not prove significantly more valuable than general populations for LD mapping of common variants underlying complex disease". Another team, headed by Pui-Yan Kwok of Washington University in St Louis, Missouri, conducted a similar study of the X chromosome using genetic markers called single nucleotide polymorphisms (SNPs) and came to the same conclusion. Both teams stress that their findings cannot necessarily be applied to the entire genome. "It shows the urgent need for much, much more data," says Kwok.

Population counts

Kári Stefánsson, chief executive officer of deCODE, says that it is still too early to tell whether homogeneous or heterogeneous populations will prove superior. "The proof of the pudding is in the eating," he says. Stefánsson adds that Iceland's meticulous medical records and extensive genealogical data, the latter of which extend back for some 1,000 years, represent a unique resource that will stand deCODE in good stead even if homogeneity proves not to be crucial. By tapping into this information, the company will be able to select healthy controls that are well matched to people affected by a given disease.

But the results from Kwok and Todd are adding to a growing impression that genetically isolated populations will not be a panacea. In Britain, for example, the Medical



Sorted: researchers prepare to genotype samples.

Research Council (MRC) and the Wellcome Trust are designing a study of some 500,000 middle-aged people. Blood samples for DNA analysis, together with lifestyle information, will be collected and correlated with the onset of diseases including cancer, diabetes and heart disease. "Our heterogeneous population, with its large ethnic minority groups, will be an advantage," says Tom Meade of the MRC's Epidemiology and Medical Care Unit at Queen Mary and Westfield College in London, who chairs the expert working group designing the study. "Our results will be representative of the population as a whole."

Researchers in Estonia, meanwhile, are preparing to launch a similar study that aims to examine one million people — three-quarters of the country's population. Scientists behind the plan had originally argued that the country's relatively homogeneous population — "somewhere between Manhattan and Iceland", according to Andres Metspalu of the University of Tartu, one of the project's organizers — would make the study easier. "But we don't think this will be our main strength now," he says. Instead, the project's organizers are putting their faith in the use of health questionnaires, carefully designed to determine who is affected by which disease, coupled with the sheer size of their sample.

Big is better

Indeed, size is the fundamental advantage offered by the heterogeneous populations found in most of Western Europe and North America. The world's largest study of cancers is the multi-centre European Prospective Investigation into Cancer and Nutrition (EPIC), which started recruiting in the early 1990s and now has 500,000 participants. EPIC's principal investigator, Elio Riboli at the International Agency for Research on Cancer (IARC) in Lyons, was wise enough to insist that blood samples be gathered from participants from the beginning. This has allowed the project to move into the study of genetics, in addition to lifestyle factors, and it has just started genotyping. So far, its sample includes 350 cases of colon cancer and nearly 2,000 of breast cancer.



Hidden gems: identifying genes that cause common diseases will mean screening the entire genome.

In the United States, the National Cancer Institute in Bethesda, Maryland, has commissioned its director of epidemiology and biostatistics, Robert Hoover, to investigate the feasibility of organizing a consortium of large-scale population studies in cancer. "Large" means all studies which are likely to generate 1,000 to 1,500 particular cancers in a reasonable time," says Hoover. This could include up to 15 studies over the next decade. Hoover is discussing whether genotyping for these projects should, and could, be coordinated, perhaps even using a single high-throughput facility. He is also considering how clinical and lifestyle information should be gathered to allow for comparisons between the different studies.

For most of the proposed work, researchers intend to focus on SNPs. These are single-base substitutions found scattered throughout the genome. They account for most of the genetic variability that helps make us all different. Large-scale identification of SNPs began only recently⁷. But since April last year, thanks to the efforts of the SNP Consortium — a collaboration funded by the Wellcome Trust together with many of the world's biggest drugs companies, and involving leading academic genomics centres — some 300,000 SNPs have already been put into a public database. That figure could top two million by next summer.

But association studies that use SNPs as markers and then fish randomly for genes that predispose to common diseases could be prohibitively time-consuming and expensive. The number of SNPs required depends on the degree of LD in the population. A thorough study could need hundreds of thousands of different SNPs from tens, or even hundreds, of thousands of individuals, sending the total number of individual SNPs to be genotyped into the billions. So far, even the biggest labs can only genotype several thousand SNPs per day. Indeed, their throughput would need to increase by some three orders of magnitude to make such mammoth studies feasible.

Prime candidates

That makes some geneticists argue that a candidate-gene approach might be the best option, at least for now. Some SNPs appear within genes of known function, and for some of these, there are already reasons to suppose that mutations in the genes might predispose people to particular diseases. For example, one study of breast and prostate cancers, being done under the EPIC umbrella, is analysing SNPs in genes involved in the synthesis of steroid hormones, which frequently influence the initiation or growth of such tumours. The research team, led by Federico Canzian at IARC, will analyse an average of five SNPs in each of 20 genes for a sample of some 1,000 cancer cases, and an equal number of matched controls. That gives a manageable total of 200,000 SNPs to be genotyped.

Meanwhile, at the French National Centre



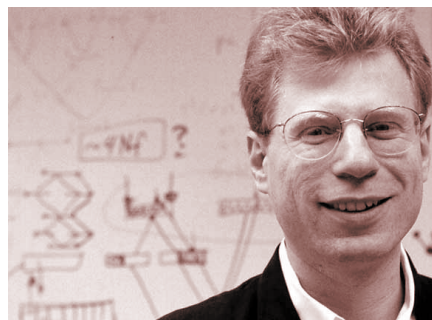
Stefánsson: deCODE controls a unique resource.

for Genotyping in Evry, near Paris, researchers led by Mark Lathrop are embarking on a larger study. They are sequencing 500 candidate genes for heart disease in 100 individuals to identify SNPs that can later be used in a study of about 8,000 individuals.

But many genes linked to disease are likely to evade candidate-gene studies, so large association studies could still be necessary. "The candidate-gene approach is the only practical way at the moment," says Hoover. "But eventually we may want to do the whole-genome association studies with no *a priori* hypothesis." Geneticists hope that advances in genotyping technology will come to the rescue, increasing throughput and decreasing costs to the point that this becomes feasible.

At present, SNPs are genotyped by two main methods. In the 'chip' approach, fluorescently labelled 'complementary' DNA sequences that bind to particular SNPs are attached to a solid surface, usually a glass slide, which is then exposed to the sample DNA. The SNPs in the sample are then identified from the resulting fluorescence pattern on the DNA chip. Once the chips have been made, this method is very efficient as it allows large numbers of SNPs to be analysed in one go. But manufacturing and processing the chips is time-consuming, and adding new SNPs to the analysis means designing new chips.

The alternative is to use mass spectrometry. The sample DNA is treated so that the nucleotides in certain SNPs are substituted



Kruglyak: doubts the advantages of isolation.

for nucleotides bearing additional chemical groups. The SNPs are then detected by analysing the molecular mass of fragments of the DNA. Only a small number of different SNPs can be genotyped at a time, but the analysis is extremely fast, and new SNPs can be incorporated into a study very easily.

Labs around the world are now scaling up both of these technologies, and working to make them more efficient. At the same time, small biotech companies are offering alternative methods, using innovative enzyme-based assays. But none of these technologies seems likely to provide the huge leap in throughput needed to conduct the thorough, genome-wide association studies that some gene hunters would like to launch.

Variety shows

Even without major technological advances, clever study design might help to reduce the amount of genotyping to a more reasonable level. Here, the focus would be on finding out more about variability in LD across the genome. Preliminary studies of chromosome 22, recently sequenced by a team led by Ian Dunham of the Sanger Centre at Hinxton, near Cambridge⁸, suggest that this variability will be high. In genomic regions where LD is extremely low, there may be little point analysing SNPs, as associations with disease will be extremely difficult to detect. And in regions where LD is very high, geneticists might be able to spot associations using a relatively low density of SNPs.

Dunham suggests that geneticists would benefit from maps detailing variability in LD across the genome for different populations. They could then decide which SNPs to use on a study-by-study basis, to ensure the best balance between cost and the efficient capture of disease genes. Kwok agrees: "It would not be difficult, and LD maps could reduce the number of SNPs needed for a whole-genome association study to as low as 30,000." Kwok is talking with other researchers about the possibility of creating such maps. "As yet there is no assured funding," he says.

As the technology now stands, 30,000 SNPs would still make association studies involving thousands, or tens of thousands, of individuals extremely daunting. But gene hunters say that such studies lie within the grasp of foreseeable developments in genotyping. "The speed of technological advance, with the resultant cost reductions, are likely to make studies on this scale feasible in the next two years," concludes David Bentley, head of human genetics at the Sanger Centre.

Alison Abbott is Nature's Senior European Correspondent.

1. de la Chapelle, A. *J. Med. Genet.* **30**, 857–865 (1993).
2. Risch, N. & Merikangas, K. *Science* **273**, 1516–1517 (1996).
3. Risch, N. *Nature* **405**, 847–856 (2000).
4. Kruglyak, L. *Nature Genet.* **22**, 139–144 (1999).
5. Eaves, L. A. *et al. Nature Genet.* **25**, 320–323 (2000).
6. Taillon-Miller, P. *et al. Nature Genet.* **25**, 324–328 (2000).
7. Wang, D. G. *et al. Science* **280**, 1077–1082 (1998).
8. Dunham, I. *et al. Nature* **402**, 489–495 (1999).

How new technology put a coelacanth among the heirs of Piltdown Man

Sir—Biology has endured a long history of fakes, ranging from Piltdown Man to feathered dinosaurs. Often meticulously crafted, these frauds were not always easy to detect.

Fortunately, technology has provided an arsenal of forensic tools that help reveal such fakery. But now digital imagery and powerful, simple-to-use photo-editing software present a new opportunity for scientific forgery. In the hands of a skilled user, this software can be used to produce almost imperceptibly altered fake photographs¹.

Fortunately, the Séret, Pouyaud and Serre image² (purporting to show a coelacanth captured in the Bay of Pangandaran off southern Java in 1995) does not fall into this category. It is clearly an altered copy of a photograph of a live coelacanth taken by M.V.E. on 30 July 1998 off Manado Tua and printed in *Nature* soon afterwards³.

While we would not be surprised if coelacanths turned up elsewhere in Indonesian waters, Pangandaran does not seem to be a likely location. Unlike the steep volcanic slopes of Manado Tua and the Comoro Islands, where populations of coelacanths are known, Pangandaran is a shallow, muddy bay.

Interviews with local Pangandaran fishermen and both the current and the 1995 chief officers of the Pangandaran Fisheries Department, by M.V.E. and scientists from the Indonesian Institute of Sciences, revealed no evidence of past coelacanth catches and no familiarity with the fish. Neither were the fishermen or fish vendors familiar with two common 'indicator species' of potential coelacanth habitat: deep-water snappers (*Etelis* spp.) or the oilfish (*Ruvettus pretiosus*).

This is not surprising, as the fisheries in the Pangandaran area consist primarily of shallow, level-bottom trawlers and open-water pelagic purse seines and longlines, with no evidence of the deep-water fishing gear that typically results in coelacanth bycatches.

Much has been made of the nationalistic animosity that has tainted the saga of the coelacanth. In our opinion, nationalism plays no role in good science, and is irrelevant in chance discoveries such as that of the Indonesian coelacanth. As scientists it is our responsibility to study and conserve. The Indonesians and Comorans are rightfully proud of efforts in their two countries to preserve these rare and very special fish. What

pride can we in the western scientific community take in this affair?

M. V. Erdmann, R. L. Caldwell

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA

1. Hayden, J. E. *J. Biocommunication* **27**, 11–19 (2000).
2. McCabe, H. & Wright, J. *Nature* **406**, 114 (2000).
3. Erdmann, M. V., Caldwell, R. L. & Moosa, M. K. *Nature* **395**, 335 (1998).

Assessors' odd listings don't inspire confidence

Sir—The assessment of research in Spanish universities, and the (lack of) consideration of scientific qualifications in appointing tenured faculty members have been the subject of commentary in *Nature* on several occasions^{1–8}. Commendably, the Catalan government has now published a proposal for assessing individual research performance in social sciences and humanities (www.gencat.es/dursi/sisav.htm). This has been drafted by a panel of experts and faculty members, who have been holding meetings since 1997.

The proposal includes a four-category ranking of journals in psychology and related fields (www.gencat.es/dursi/sisav_i_12.htm). The top category reportedly includes top-quality international journals with the highest impact indices in their fields, whereas the bottom category includes journals lacking a rigorous peer-review process.

Notably missing are journals such as *Nature*, *Science*, *Vision Research*, *Visual Neuroscience*, *Spatial Vision* or the *Proceedings of the National Academy of Sciences*, to name but a few — although the top category includes the hitherto-unknown "Nature Neuropsychology". Also in the top category is "Journal of Experimental Psychology Human" while the bottom category lists "Journal Experimental Perception Performance". One wonders whether the genuine *Journal of Experimental Psychology: Human Perception and Performance* would have fallen into the top or bottom category if it did appear.

Again, the top category contains "Journal of Experimental Psychology Learning", whereas "Journal Experimental Psychology Learning Memory Cognition" only makes it into the bottom category. Where would *Journal of Experimental Psychology: Learning, Memory and Cognition* have appeared?

The second-best category includes "Applied Psychology Measurement", whereas the real *Applied Psychological Measurement* is in the bottom category. There are myriad cases of this type.

As a psychologist who has always wished for research performance to be judged on the basis of well-pondered information, I can only hope that the ultimate ranking of journals will be made by strict adherence to

relevant criteria, and by a panel that shows evidence of knowing what journals exist.

Miguel A. García-Pérez

Departamento de Metodología, Facultad de Psicología, Universidad Complutense, Campus de Somosaguas, 28223 Madrid, Spain

1. *Nature* **396**, 709 (1998).
2. Bosch, X. *Nature* **396**, 712 (1998).
3. Escartin, J. *Nature* **401**, 112 (1999).
4. Bosch, X. *Nature* **401**, 419 (1999).
5. Bosch, X. *Nature* **402**, 848 (1999).
6. *Nature* **404**, 211 (2000).
7. Mira-Pérez, J. *Nature* **404**, 222 (2000).
8. Bosch, X. *Nature* **404**, 425 (2000).

Reductionism should be clarified, not dismissed

Sir—John Ziman, reviewing Norman Levitt's book *Prometheus Bedeviled* (*Nature* **404**, 811; 2000) calls reductionism a "busted flush". But what is reductionism? Molecular biologists in particular tend to be accused of it, but they do not hold the naive view that complex structures and processes are just sums of their parts, with the implication that one can neglect what are sometimes called 'emergent properties'. Biochemists and molecular biologists are now preoccupied with macromolecular complexes, trying with some success to explain how their amazing activities emerge from molecular interactions.

What the so-called biological reductionists believe is that it is possible in principle to explain (not usually predict) the interactions within complex systems in terms of the universal principles of chemistry and physics. There is nothing to show that this kind of reductionism is wrong.

It does not imply that the molecular analysis is going to be easy. Evolution has produced systems of infernal complexity, and molecular biologists analyse them as best they can, in the belief that even partial knowledge is worth having. Of course, for many practical purposes one has to rely on generalizations at more superficial or, if one prefers, higher levels, leaving deeper analysis for the future. Some people hope for the emergence of new laws of physics that will simplify complex systems, although it is difficult for most of us to believe that this will ever happen. The other means of escape from the hard work of molecular analysis is vitalism, nowadays more implicit than overt: the feeling that the molecules of life must be imbued with some essence that science cannot reach.

Arguments about reductionism would be more valuable if there were greater clarity about what it is, and what its critics would like to have instead.

J. R. S. Fincham

Institute of Cell and Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, Scotland

Pitching it right

Observing chimps and the women who observe chimps.

Africa in My Blood: An Autobiography in Letters

by Jane Goodall
Houghton Mifflin: 2000. 371 pp. \$28

Beauty and the Beasts: Woman, Ape and Evolution

by Carole Jahme
Virago: 2000. 406 pp. £18.99

Alison Jolly

Jane Goodall produces one extraordinary book after another. This one she never meant to be published, and it is one of the best of all. Her early letters fizz with the excitement of a young woman making her first discoveries. Because that woman is Jane Goodall, her discoveries changed our view of human nature, not just chimpanzees. The letters convey it all as breaking news, while Dale Peterson's superb editing and linking introductions keep up the pace.

Some readers may yawn at the early outpourings of a horsey, rather bossy little English girl. Others who have raised, or been, such children will grin in recognition, then wonder how it was that Goodall emerged from the stereotype. Most readers will then be astonished at the giddy whirl of an unattached blonde in Kenya Colony on the eve of independence. Also at the frankness of her letters home: "It is disconcerting being proposed to every ten minutes ... [and] what the devil am I to do with all these middle aged married men. They hang in multitudinous garlands from every limb and neck I've got."

When Goodall reached the Gombe Stream Chimpanzee Reserve, there was the long bout of malaria, then the slow habituation of chimpanzees. Her first view of the 'Rain Dance' was across a valley. She could



Chimp appeal: Goodall recognized her subjects' individuality and emotional complexity from the start.

only just identify sex and face colour. It gave a sweeping panorama, though, of the males as they "hurled down [the hill], one after the other, leapt into a tree, seized branches and leapt to the ground — 30 ft at least, taking their branches with them, charging on, reaching a tree, seizing it, swinging round and leaping up into the branches ... the women and children watched in silence, and the rain poured down while the lightning flashed brilliantly across the grey sky ... Can you begin to imagine how I felt? The only human ever to have witnessed such a display, in all its primitive, fantastic wonder? I fear I was not able to report it exactly as science might wish. I had no thought for exact numbers, or seconds — nor could I write as I watched — a) for fear of missing one second, and b) because it was far too wet."

To my mind, Goodall has been blessed with a kind of perfect pitch. She never yielded to the urgings of journalists to camp up her animals, or lessened her respect for the creatures she studies by imputing human attitudes they probably do not have. On the other hand, she never gave in to some scientists' insistence that animals should be made duller than they really are. Goodall recognized her chimpanzees' individuality and emotional complexity right from the start, and has shifted Western science's approach to animals as a consequence.

Academics who wish to trace the philosophical bases of anthropomorphism as a threat to, and as a tool of, science will find these letters a rich vein of source material. The over-intimate contact of the banana-feeding days, the compromises necessary to

get good photos for the *National Geographic*, then the drawing back, for the safety of human assistants and the chimpanzees' own freedom, are all here. By the end of the book, Goodall is a published author, beginning the jet-set life of dedication which has taken her so far from the Gombe Stream.

I wish that Carole Jahme shared the same perfect pitch. She has a goal: to show that women love wild primates more than men do. Women, she says, see monkeys and apes as individuals whose fascination may outrank those of lovers, husbands, or the hope of children. Men, in her view, spend just long enough in the field to gather thesis data, then rush home to build a career. "Established male scientists undertake sporadic trips to research centers, but during these short visits it is rare for them individually to gather enough data to substantiate a scientific paper, let alone fill a book." I can see right now a few of my colleagues' beards and neck-hairs beginning to bristle.

Overstatement, as all journalists and most academics are aware, is the way to sell copy. The problem is that Jahme is half right. Women scientists have indeed devoted their lives to specific animals and study sites. Many even dare to champion emotion as well as intellect. Jahme quotes Sue Savage-Rumbaugh: "I wonder if Piaget had been female if he would have been accused of finding these things in his children because he was their mother? I think there is a bias against females. It's okay for males to have empathy and be objective, but females are presumed to have some kind of difficulty in being empathetic and being objective at the same time. I



Back on home territory: travelling to a London TV programme with Annabel the orang-utan.

book reviews

think that's a false dichotomy." She quotes, but seems far less at home with, Sarah Hrdy's flat statement: "I did not come into science and primatology because of a love for non-human primates. I was first attracted to problems and to things I wanted to understand ... There is really no one way of doing science ... we need people to stress theory as much as observation." One of Jahme's subjects, reading the section about herself, murmured: "Oh dear. This is half right and all wrong." To me that sums up the book, including some amazing scientific slip-ups.

A chapter I did like treats apes in the media. Jiggs, the chimp that played Cheeta in the *Tarzan* films, actually loathed women, including "Jane", whom he apparently cosies up to on the book's cover. (The dislike was mutual.) More telling is an incident when three television crews simultaneously followed Birute Galdikas through the forest, filming over each others' shoulders. The Japanese director insisted that Galdikas share a cup of sugary tea with a rehabilitant orang-utan. Jahme and her own crew winced, knowing how their British audience idealizes "apes in the wild". Galdikas humoured the Japanese. Her priority was to do whatever she could to influence a country whose loggers were felling Bornean rain-forest. People are willing to sell their image when it gives the research and the animals themselves more chance of survival.

That same dilemma appears in detail in Goodall's letters, with the nuances that inevitably escape Jahme's summary. You can read *Africa in my Blood* as a key document in the relation of human and animal, and of science and the demands of fame. Alternatively, just read it for fun. ■

Alison Jolly is in the Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544-1003, USA.

The spice of life

The Variety of Life: A Survey and a Celebration of All the Creatures That Have Ever Lived

by Colin Tudge

Oxford University Press: 2000. 684 pp.

£35, \$49.95

Pat Willmer

There is never a good time to write a book, especially one that addresses issues currently exercising the ingenuity and the advancing technology of biologists. This volume sets out to give us an overview of all living organisms and how they are related, and then doubles the difficulty of the task by including all the known extinct life-forms as well. At the turn of a century that has seen this field become the focus of an acceler-

ating input from new methodologies, this is heroism. Since the author took 10 years over the job, he must have been daunted by how much the ground changed beneath his feet as he progressed. When the world around you is bursting with new kinds of information that are just beginning to shed fresh light on a key question, it is a brave man who decides to write about the answer! So why has Colin Tudge done it?

Tudge's answer is twofold: to help put taxonomy back at the centre of biology (and biodiversity) teaching, and to point out that nature is wonderfully varied. He aims to help professionals knowledgeable about one kingdom to find out about the others, and to help amateurs find their way around all the kingdoms (especially the animal kingdom, and it would help — as always — if they are particularly keen on vertebrates!).

The first four chapters do a great job of acquainting non-biologists with the nature and history of taxonomy and the impacts of cladistic and molecular analyses; they should be widely read (although many will dispute Tudge's emphases). Thereafter, however, non-biologists may find themselves overwhelmed by detail.

Professionals will probably pass over these first chapters and go for the detail, and they probably *will* go for it — in a field such

as this there is much to take issue with. In my own areas of special interest I found almost more to question and argue with than to agree with; and Tudge inevitably can only afford space for a thin overview of the real issues that excite us today. Indeed, for animal phylogeny, the book only sparsely applies the information that is now emerging from 'evolutionary developmental biology' (which is revealing underlying similarities between diverse species in the structure of the genes that control pattern formation and organ and body shape); its implications pass largely unnoticed.

The author sensibly gets his retaliation in first by telling us that he knows he will be criticized for falling between two stools — but he does, and it was inevitable that he would. Despite that, this is a lovely and accessible book, fun to dip into and fun to disagree with. It will be hugely valuable as a source-book for student libraries, and highly informative for any enthusiastic (and patient!) lay naturalist curious about the life around them and the fossils beneath them. There is also a fund of diverting notes on the more curious of surviving and defunct living things.

But I wish the book were either a bit longer and more colourfully illustrated, or rather shorter, so that it would be able to



Snakes in the grass ... and elsewhere

The feathered bush viper shown above is to be found in the trees of equatorial Africa, whereas the coral cobra on the left inhabits the rocky outcrops of southern Africa's dry valley plains. The lifestyles, diversity, evolution and biogeography of these and many other variants of the predatory limbless lizard are addressed in *Snakes: The Evolution of Mystery in Nature* by Harry W. Greene, with photographs by Michael and Patricia Fogden (University of California Press, \$29.95, £18.95; pbk).

compete better in a difficult market-place. It deserves success for its heroic ambition. ■
Pat Willmer is in the Bute Building, School of Biology, University of St Andrews, St Andrews KY16 9TS, UK.

Tangled strands in the double helix

It Ain't Necessarily So: The Dream of the Human Genome and Other Confusions

by Richard Lewontin
New York Review of Books/Granta: 2000.
 330 pp. \$24.95/£14.99

The Triple Helix: Gene, Organism, and Environment

by Richard Lewontin
Harvard University Press: 2000. 144 pp.
 \$22.95, £14.50

Mark Ridley

Richard Lewontin is well known for his many positive contributions to evolutionary genetics, particularly the use of gel electrophoresis to measure genetic variation. He is also well known for his negative criticisms of certain genetic (or apparently genetic) ideas about humans, and of the 'reductionist' biological thinking that he believes underlies those ideas.

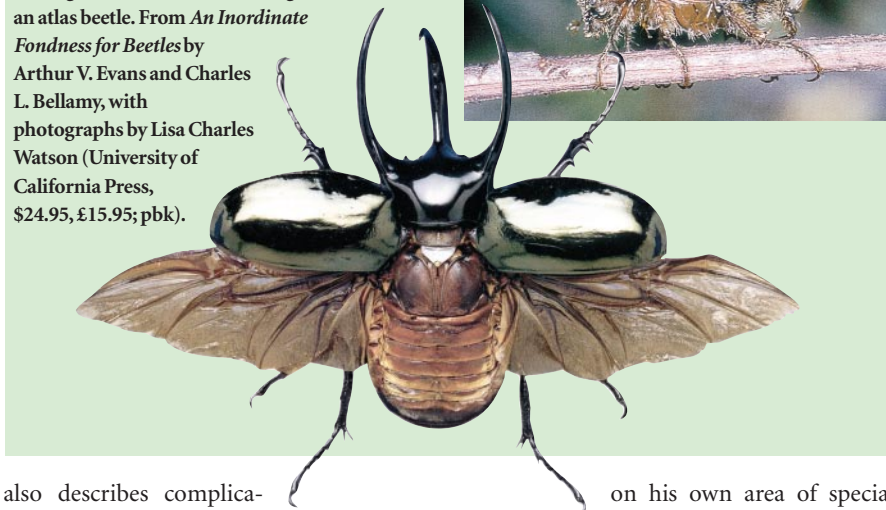
These two books are mainly in the second, critical mode. *It Ain't Necessarily So* is a collection of nine pieces from *The New York Review of Books*, together with short updating epilogues; they are mainly about humans. *The Triple Helix* is more abstract. It is based on three lectures that were published in Italian in 1998; Lewontin has added a fourth chapter on "New directions in biology" for the English edition. *The Triple Helix* is written more for biologists, although it assumes little technical knowledge: it is the material, rather than the level of treatment, that will (I suspect) mainly confine it to biological readers.

The Triple Helix looks at two big biological questions: development and adaptation. On development, Lewontin takes as text Sydney Brenner's remark that, if he had the complete sequence of DNA of an organism and a large enough computer, he could compute the organism. Lewontin describes this as "bad biology", because of the influence of the environment and random factors in development. On adaptation, he criticizes the idea that the external environment poses problems to which the organisms evolve solutions. This, too, is "bad biology" because the organism influences its environment too, through behavioural choices as well as by actively modifying it. He suggests that a better metaphor would be to say that the organism constructs its environment. He

A case of beetlephilia

Mating melolonthine scarabs (right) and an atlas beetle. From *An Inordinate Fondness for Beetles* by

Arthur V. Evans and Charles L. Bellamy, with photographs by Lisa Charles Watson (University of California Press, \$24.95, £15.95; pbk).



also describes complications caused by limited genetic variation, multiple peaks (one or two horns on a rhino) and epistatic fitnesses (in the Australian grasshopper *Moraba*).

These points, as Lewontin says, are well known to biologists. "But this in-principle knowledge cannot become folded into the structure of biological explanation unless it can be incorporated into the actual work of biologists," he adds.

This is the point at which the real controversy begins. In the case of adaptation, biologists have made more progress than Lewontin recognizes. Organisms do indeed influence, and in a sense partially construct, their environments. But methods exist to understand the situation, at least for frequency-dependent selection. In social behaviour, the fitness of an organism depends on what other organisms are doing; each organism's behaviour influences its social environment. The sex ratio is another (non-behavioural) well-analysed case. The organisms and their social environment "coevolve", as Lewontin says.

But what he does not say is that game theory allows us to model what will happen. Lewontin takes a negative view of what he calls "the metaphor of adaptation", saying that it "is now an impediment to a real understanding of the evolutionary process and needs to be replaced". But there are strands, such as game theory, within existing research that should contribute to the constructionist research programme he foresees.

Even for readers who do not agree with Lewontin, there is much of value in both books. He is superb at conceptually characterizing large research programmes in biology, and putting them in their historical context. In the penultimate paragraph of *The Triple Helix*, for instance, he comments

on his own area of special expertise. "Before the mid-1960s, experimental evolutionary genetics included a wide diversity of questions," he says, going on to list them. One of these questions was answered by gel electrophoresis. "The result was an abandonment of the research in almost all aspects of evolutionary genetics other than the characterization of genetic diversity. A single, easily acquired technique changed and pauperized, temporarily it is to be hoped, an entire field of study."

His writing is consistently elegant and readable, frequently funny, and abounding with provocative remarks. He can be a ferocious polemicist, and an effective (if sometimes slippery) debater — I'd trust him always to say something interesting, but not always to be fair to an opponent. One of his pet hates seems to be a remark made by Richard Dawkins [in *The Selfish Gene*] that the genes have "created us, body and mind". *New York Review* readers should have got the point, because Lewontin quotes it in four of the nine reprinted reviews. What he does not tell them is that the word "created" has an evolutionary, not a developmental, sense and does not illustrate the "genetic determinism" (or, more accurately, genetic fatalism) that Lewontin is mainly concerned to criticize.

The publisher George Weidenfeld once received a phone call from Thomas Wilson, head of Harvard University Press, saying, "I've got something important for you. Please don't ask any questions, just come over." "Next morning," Weidenfeld says, "I flew to Boston and called on Wilson. He closed the door and said in an emphatic stage whisper: 'the syndics of Harvard Press have seen fit to reject a manuscript of epochal importance.'" It was *The Double Helix*. J. D. Watson ("the great panjandrum of DNA") predictably gets the sharp end of Lewontin's

pen. I do not think that Harvard University Press has a book of epochal importance in *The Triple Helix*. Most of the points, and the examples Lewontin uses to make them, will already be familiar to biologists. He has used the same examples in Suzuki's genetics text (of which he is a co-author) *An Introduction to Genetic Analysis* (Freeman), in anti-adaptationist papers and in some other books.

However, both these new books could provide an easy way for biological readers to refresh their knowledge of, or argument with, Lewontin's views. The books' main value will probably be for biology students who have not yet met those views. And I cannot recommend them too highly for the many commentators and headline-writers who think that DNA is the blueprint for the organism. ■

Mark Ridley is in the Department of Zoology, University of Oxford, Oxford OX1 3RS, UK.



Founding figure: Poincaré's visionary work opened up the huge subject of topology to investigation.

Story of a favourite mathematical tool

History of Topology

edited by I. M. James

Elsevier: 2000. 1,056 pp. £125, \$190.50

Jeremy Gray

As topology may well claim to be the twentieth century's most original and successful development in mathematics, a comprehensive history of it is most welcome. In its algebraic form, on which this volume focuses, topology provides a way of formulating and often answering questions about the geometry of spaces in any number of dimensions. But it reaches further than that, and, like any major branch of mathematics, has its own charm and its own difficulties, attracting some of the world's best mathematicians.

Topology owes its range and depth to the primitive nature of the mathematical features it studies. These features are preserved even when a figure continuously changes its size or shape, including bending and stretching, or even undergoes certain kinds of cutting and re-glueing. So whenever a property is shown to be topological in nature, even though it might seem to need more sophisticated concepts for its definition, mathematicians infer that they have got closer to the heart of the matter. Two classic examples of this are the number of independent integrals on a surface and the study of differential equations on a surface. The nineteenth-century mathematicians Bernhard Riemann and Henri Poincaré showed that these calculus-type properties of a surface depend on a single number, called the genus of the surface, which is a true topological invariant.

The book's editor, I. M. James, himself a distinguished topologist, has drawn together

more than 40 authors for this account. Some are eminent mathematicians, others are historians of mathematics. Each group displays many of the skills of the other, so the reader enjoys a consistently rich but varied diet.

Inevitably, and rightly, some essays are for mathematicians only. But it is possible to find others that show how topology acquired its present central position. A founding figure is the French mathematician and scientist Henri Poincaré. His visionary but often inaccurate work is well described by Karanbir Sarkaria, who is a careful and well-informed guide. Poincaré's achievement was in opening up a huge subject to investigation. As Erhard Scholz nicely describes, the theory of manifolds (generalizations of smooth surfaces to any number of dimensions) links Poincaré's work to the more general study of the calculus of several variables. Such generalization goes back before Poincaré to Riemann, but Poincaré promoted it, even though he left it to others to find the right tools for the job. And there is a fine essay by Moritz Epple showing how problems in knot theory, which is rooted in nineteenth-century physics, led to the creation of rigorous methods in the study of three-dimensional manifolds.

Some of the spectacular applications of topology are also discussed. The study of Lie groups may have been started by the Norwegian mathematician Sophus Lie, after whom they are named, but, as Thomas Hawkins lucidly describes, the modern formulation owes much to Hermann Weyl, who showed how crucial topology is to that subject. Currently, topology is a favoured topic among physicists, and Charles Nash's essay moves quickly over the period covered by Epple to describe the exciting developments

in such areas as mirror symmetry, black holes and gravitational collapse.

Algebraic topology is a sophisticated machine that students find hard to master, and many of its main tools are considered here. There is a fine account by John McCleary of the early history of spectral sequences, one of topology's most complicated aspects, and another by M. Zisman on fibre bundles (loosely, parametrized families of groups or manifolds).

The book contains a welcome emphasis of some 180 pages on topologists themselves, either singly or in schools. Much of the information here is new, and opens up fresh perspectives on some of the material described in earlier essays in the volume. The handy 24-page index is another bonus.

It is no criticism of a book of more than 1,000 pages to say that much has been left out. There is nothing here on the Russian school of topology, in which Lev Semionovich Pontrjagin and Pavel Sergeevich Alexandrov figured so prominently. Many branches of physics draw on the methods and ideas of algebraic topology, but some of these are perhaps too recent to be included in a history. Surprisingly, analysis and elliptic operators are mentioned only in passing. This is a pity, for the subject reaches from the work of Riemann to the Atiyah–Singer index theorem and modern quantum field theory. Faced with such a huge story, however, James is to be thanked for organizing so much information and solving the editorial problems so well. This book is not meant to be the last word on anything. It does much to enable the next words to be said. ■

Jeremy Gray is in the Department of Mathematics, Open University, Milton Keynes MK7 6AA, UK.

What if...?

...Chinese explorers — or Darwin's bulldog — had settled in Australia?

John Carmody

If, as John Lennon sang, life is what happens to us while we are busy making other plans, then science is no less replete with 'might have beens'. For those of us who work in the Antipodes, but recognize that the centre of scientific and economic gravity is in the north, there is a particular piquancy about the contemplation of intellectual and cultural 'what ifs'.

James Cook made the determining discovery of Australia in 1770 but, with only sporadic contact by previous explorers, the discoverer might just as easily have been Spanish, Dutch, Portuguese or French. In fact, a Shou Lao statuette, discovered near Darwin in 1879, is possibly a relic of a visit by the great Chinese explorer Ch'eng Ho in around 1430, during the reign of Emperor Yung Lo. But the emperor died soon after this, and, owing to courtiers' opposition, Ho's voyages ceased — leaving Australia for European discovery. How would science have fared if Chinese policies had been different — given that, to quote Joseph Needham, "Chinese civilisation could not of itself produce modern natural science"?

Cook came to Australia on behalf of the British government and the Royal Society, leading a voyage with an important scientific mission — to observe the transit of Venus from the South Seas. He also carried a secret commission to determine the existence of a southern continent postulated by geographical philosophers. The presence on Cook's ship, the *Endeavour*, of the young Joseph Banks and Daniel Solander highlights that scientific purpose. The flora and fauna that they brought back to London were a crucial challenge to scientific orthodoxy, and would later aid the flourishing of Darwinism.

Darwin's own intellectual achievements had their genesis in the Pacific when he was aboard the *Beagle*, which visited several places in the fledgling Australian colony in 1836. Although influenced by what he observed of the plants, animals and aboriginal people, there was little chance that Darwin would stay in Australia. But it was otherwise for T. H. Huxley, who arrived in 1847 as surgeon-naturalist on the *Rattlesnake* and, with the support of the distinguished Sydney naturalist W. S. Macleay, described numerous marine invertebrates collected during the voyage. Huxley's paper "On the Anatomy and Affinities of the Family of the Medusae" (1849), read to the Royal Society on his behalf, secured his fellowship at the age of 26. But what if, in 1850, the nascent University of

Sydney had not, for undocumented reasons, abandoned plans to include a foundation chair in natural philosophy? Being both distinguished and on the spot, Huxley would have been the natural choice for this position. "Had the University of Sydney been carried out as originally proposed," he later wrote to Macleay, "I should certainly have been a candidate for the Natural History Chair. I know no finer field for exertion for any Naturalist than Sydney Harbour itself."

Had he remained, Huxley's influence in this paradise of biodiversity would have been incalculable — but what might this have meant for the acceptance of Darwin's theory of evolution and natural selection? Huxley's confrontation with Bishop Wilberforce is legendary, but his importance in the radicalization of European scientific thought cannot be overstated.

What, too, if the Braggs had remained in Australia? Sir William did important work on radiation in Adelaide, but the experiments for which he and his son, Sir Lawrence, won the Nobel prize were performed in England. And if Howard Florey had returned to Adelaide instead of remaining in Oxford, what might have been the story of penicillin?

And what if clinicians at Sydney Hospital had been more understanding towards the work done there in the 1940s by the remarkable trio of John Eccles, Bernard Katz and Stephen Kuffler? Katz and Kuffler were

refugees from Hitler, but the idyllic period of their collaboration with Eccles on problems of neuromuscular transmission could not survive the animus of those who thought that making plasma for troops at the front was far more important. All three left Australia prematurely, Eccles going to a chair in Dunedin, New Zealand, before moving to Canberra. Two of this trio went on to win Nobel prizes, and Kuffler is surely one of the greatest scientists not to have won one.

While in Dunedin, Eccles met Karl Popper, another scientist effectively rejected by Sydney University for xenophobic reasons, whose influence on him was enormous. Eccles came to understand that discarding his dogged adherence to the notion of electrical transmission at synapses was not intellectual death but, rather, that "killing our theories by superior ones" is essential for intellectual survival. What if this pair had met and worked in Sydney?

"You should have been here last week" is almost a proverb in Australia. Every event always happened 'last week' in a culture sceptical of ideas and prone to let ability and inventions slip through its fingers. This is part of a colonial inheritance, with its perennial fretting about the brain drain. But such might-have-beens can cut both ways, as life flows on.

John Carmody is in the Faculty of Medicine, University of New South Wales, Sydney 2052, Australia.



All aboard? If the *Rattlesnake* had left Sydney without Huxley, science might look very different today.

STATE LIBRARY OF NEW SOUTH WALES



The bottle

Force the weird stuff down.

Tom D. Schneider

He discovered that his teacher didn't know much about survival without high-tech backup systems. He almost died several times from bad food and exposure. But he learned and he survived.

One day, many years later, he found a glass vial on the shore, surprisingly unbroken, tangled in seaweed. It was only three or four centimetres long, with corks on both ends and green growth under the glass on one cork. The liquid inside was gold and glittered with purple speckles. On the glass was etched "DRINK ME".

He vaguely remembered his mother urging him to "swallow your nanos" while a doctor looked on. He had thrown a screaming tantrum and, to the horror of his mother, one vial had spilled on the floor. Finally his mother said "When you swallow your nanos we'll go get ice cream!" and he forced the weird stuff down.

He realized that the vial in his hand contained a child's level of power because it had two openable ends. This had been a fad to keep the kids intrigued because it gave them a 'choice'. It was not as powerful as he would have liked, but it would have to do.

He took the bottle to the stream, rinsed it off, then opened it and swallowed the contents. They tasted of sand and apricots.

He sat down and watched the stream calmly meet its dispersion into the sea. He imagined the gritty particles attaching to his throat and stomach. Other tiny machines migrated into his blood and dispersed

around his body, integrating themselves into his systems and awaiting his commands.

He decided to build a shelter again. The last one had been swept out to sea in a monsoon. So he had the machines put him into a dream-filled sleep.

He awoke with a start at sunset when the birds became silent. A green flash marked the last rays and the stars blossomed. In his hand was another vial. It was the same as the first, but new and the inscription read "INTO THE POND". The machines had answered his wish for shelter!

He carried the vial inward to the one fresh water pond on the island, opened it, and poured the swirling contents into the emerald waters. He ate some figs and lay down to watch the stars and the passing satellites.

Hot sun on his face awoke him. The pond was purple and sparkled with gold. For the first time in many years he grinned. Then he laughed, rolling on the sand until tears came to his eyes.

He knelt at the edge and inspected the pond, which was now an industrial fluid. Purple bacteriorhodopsin was absorbing sunlight and pumping hydrogen ions into nanotube fuel containers. The gold-silicon computer chips floating in the pond were under his direct mental radio control, but the pond needed time to absorb enough sunlight to do its work.

He walked down to the sea to catch some fish with his woven reed net. He built a fire, starting it with the ancient bow method, and cooked the fish. The carcinogens in the fried meal were no longer going to be a problem.

One evening a few days later he returned

to the pond. A quaint wooden cottage stood next to it; polished wood floors and furniture gleamed through the open door. There were wooden bookshelves and cotton sheets with a quilt on the bed. He sat on the bed and cried in loneliness.

I am male, he thought. Take a cell, remove the Y, duplicate the X. He lay on the bed and slept.

The next day he set about filling his house. He sculpted sand into glasses and wind chimes, he pulled silver, iron and gold from the sea to make utensils. At a whim, he downloaded books by growing paper on a tree with the words already printed on it, and filled the empty shelves.

He didn't dare make a boat. They would not welcome him back.

One morning a squall awoke him. It was raining lightly outside and he ran to the pond's edge. A beach ball sized egg had rolled out of the pond onto the shore and had cracked open. A baby tangled in goo cried inside. He cleaned the child in the purple glittering pond water and fed her milk from the dripping pods of a new plant that grew at the pond's edge.

But the child did not breathe well and died by the third day.

In his grief he sat by the sea and watched the bottlenose dolphins play, leaping in the surf. He longed to be with them. He walked into the sea and sat in the shallow water to wait. He slept in the hot sun. ■

Tom D. Schneider is a molecular information theorist and nanotechnologist at the National Institutes of Health in Frederick, Maryland. <http://www.lecb.ncifcrf.gov/~toms/>

How robust is the Internet?

Yuhai Tu

Complex systems, such as the Internet, are surprisingly resistant to random errors. But a new study warns against complacency — the feature that makes the Internet immune to accidents also makes it vulnerable to attack.

We are living in a connected world. Networks permeate every aspect of our lives. Our consciousness is based on the network of neurons in our brain. Our society is formed by various social, economical and political networks. The Internet, which is profoundly changing the way we live, is nothing but a network of information resources. Our overall well-being depends to a large extent on the stability and health of many networks. Yet we live in a constantly changing and often hostile environment. It is therefore important to understand how robustly these networks respond to disruptions. On page 378 of this issue, Albert, Jeong and Barabási¹ address this problem in a simple and elegant way. They find that the tolerance of networks to different types of perturbation depends critically on the network structure.

Given a set of nodes and links, a simple network can be built by linking pairs of nodes at random until we use up all the available links². Such a random network belongs to the class of 'exponential networks' because each node has roughly the same number of connections (it is statistically homogeneous) and the frequency of highly connected nodes decreases exponentially. But most naturally occurring networks have much more intricate hierarchical structures; for example, the frequency of highly connected nodes often decays as a simple power law. Power laws are found in many areas, from the distribution of population sizes to the frequency of scientific citations. Networks that follow power laws are called 'scale-free networks' because they are not tied to a specific scale. This means they are extremely inhomogeneous: whereas most nodes have one or two links, a few highly connected nodes will have a large number of links and so play a key role in the behaviour of the network.

Albert and co-workers¹ wisely simplify their study of network tolerance by focusing on these two classes of network structure: the random, exponential network and the self-organized, scale-free network. To measure the quantitative performance of a network, the authors use the average distance between two nodes in the network, where distance is defined by the minimum number of links between these two nodes. The authors also consider two types of destructive perturbation: randomly occurring error, which is due

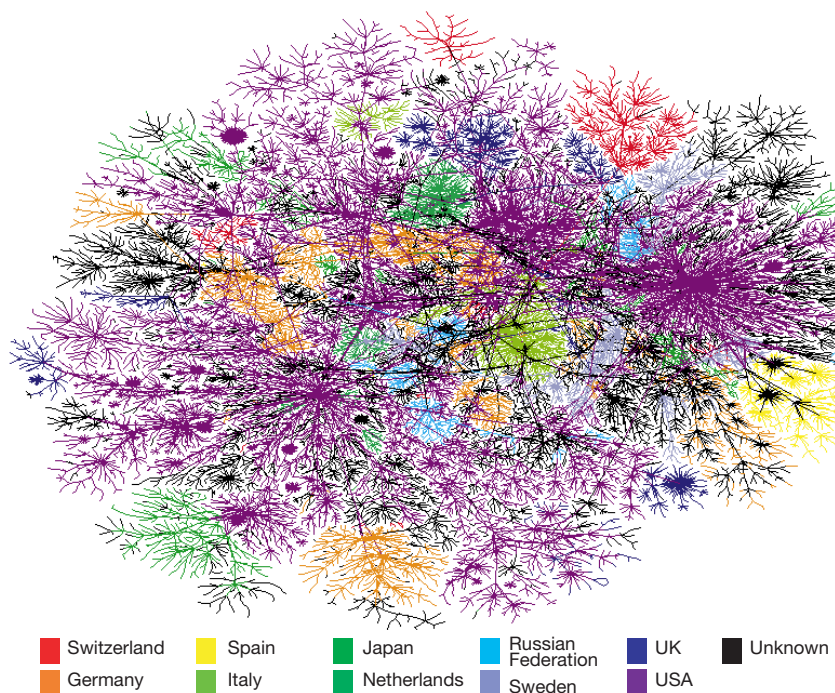


Figure 1 What does the Internet look like? No map exists of the entire Internet, but these lines show the paths an e-mail might take across some of the largest networks. The lines branch at each network router, or node, along the way. Colours were assigned according to the geographic domain (for example, .se for Sweden) where each network router was registered. The map was created using the skitter tool (developed by D. McRobb at CAIDA)⁷, which sends out small packets of data from a source to many destinations through the Internet. The data collected by skitter give a snapshot of the Internet at a particular moment. (Graph created by B. Huffaker using graph layout code provided by B. Cheswick and H. Burch.)

to random malfunction of the nodes; and an intentional attack, which is aimed at the most connected nodes.

For an exponential network, Albert and co-workers find that the accumulation of random errors has a deteriorating effect on the network performance. This decrease in network performance occurs because each deletion of a node destroys some local paths, which leads to an increase in the distance between the nodes involved in these paths. The above argument seems quite reasonable and general. But in the case of a scale-free network, Albert and co-workers make a surprising discovery: they find that the performance of the scale-free network is almost unchanged by the random removal of nodes up to a large deletion rate.

The immunity of this network's performance to random error suggests two features about the network structure: first, most of

the nodes are just 'end users', the removal of which does not affect the paths between other nodes; second, there are 'degenerate paths' between nodes, which implies the existence of highly connected nodes. These two competing requirements have apparently found the perfect balance in the scale-free network. Further study is needed to understand this network's immunity to random error — one can only speculate that the reason lies deep in the scale-free nature of the network structure.

However, the structure that makes the scale-free network superior to the exponential network in the case of random error becomes its Achilles' heel under hostile attack. The most effective way of destroying a network is to attack its most connected nodes. For the scale-free network, in which there are nodes of high connectivity, the effect of targeted attack is much more severe



100 YEARS AGO

Snake-stones are fairly common in South Africa, and are described as white, porous stones, which, when applied to the place where the snake has bitten a person, adhere till all the poison is drawn out into them, after which they are placed in milk, which in turn draws the poison from the stones, and renders them again fit for use. The farmers firmly believe they are taken from the head of a snake. It is suggested that snake-stones are made of pumice. To the uneducated, the structure of pumice has a close resemblance to that of bone, and this may possibly explain the popular delusion that snake-stones are made of bone... The fact that the fable of the stone having been taken from the head of a snake is exactly the same in the Malay States as is prevalent in South Africa is interesting, though the Malay slaves which the early Dutch obtained from Batavia in exchange for quaggas, zebras, ivory, &c., may have carried the legend with them. It is not an uncommon custom in Germany for people to carry about with them nuggets of raw gold to draw out of their bodies all the more subtle evils, such as those produced by spirits and devils, while for the grosser evils they carry a potato. Is the snake-stone legend a derivative of these, or are they subsequent to the snake-stone?

From *Nature* 26 July 1900.

50 YEARS AGO

The mango, a favourite tropical fruit, which is widely cultivated in India, belongs to the Malaysian genus *Mangifera* Linn. (fam. Anacardiaceae). Taxonomic study shows that it contains forty-one valid species, three of which, *M. indica* L. (wild and cultivated), *M. sylvatica* Roxb. (wild in the hilly forests of north-east India), and *M. khasiana* Pierre (a species of doubtful occurrence) have been reported from India. About a thousand cultivated varieties of mango occur in India, all of which are included in the single species, *M. indica* L. (including twenty-three grafted varieties, and one wild race), *M. sylvatica* Roxb. and *M. caloneura* Kz. (a Burmese wild species) — show striking stability in their chromosome numbers, all having $n = 20$ and $2n = 40$ chromosomes.

From *Nature* 29 July 1950.

than that for the exponential network, in which all nodes are statistically identical. To put their theory to test, Albert and co-workers analyse two of the most popular networks: the Internet and the World-Wide Web. Although many people use these terms interchangeably, for the purposes of this study they are considered as separate entities. The World-Wide Web is defined as the network of web pages (the nodes) joined together by hyperlinks, whereas the Internet is the physical communication network linked together by routers (Fig. 1). Albert and co-workers discover that the error tolerance of these two networks has exactly the same characteristics as that of the scale-free network. This result, together with earlier studies^{3–5} revealing the power-law structures of these networks, confirms the connection between network structure and performance. More importantly, it provides valuable insight into the structure and weakness of these essential networks.

What can we learn from this study? The good news is that we do not have to worry about random fluctuations of these networks. The bad news is that Internet terrorists could cause great damage by targeting the most connected routers or web sites. The average performance of the Internet is reduced by a factor of two if just 1% of the most connected nodes are destroyed; and with only 4% of its most important nodes destroyed, the Internet loses its integrity, becoming fragmented into small disconnected domains. Of course, modelling the real Internet means taking into account such details as the bandwidth of different links,

the varying error susceptibility of different nodes and, most importantly, different communication protocols. It remains to be seen how these detailed considerations will affect the conclusions of the current study. Nonetheless, this work represents a notable first step towards understanding the robustness of the Internet.

More generally, the work of Albert and co-workers provides a useful framework for qualitatively describing and analysing network performance. It will be interesting to see what we could learn about other complex network systems, such as neural networks or gene regulatory networks⁶, by using similar analysis. In these biological systems, error tolerance is not just a passive property of the network structure; rather, it is part of the driving force by which evolution selects the network structure. Perhaps if we could understand why certain network topology is preferred and selected by nature, such knowledge could ultimately help us design more robust artificial networks. ■

Yuhai Tu is at the IBM T. J. Watson Research Center, PO Box 218, Yorktown Heights, New York, New York 10598, USA.

e-mail: yuhai@watson.ibm.com

1. Albert, R., Jeong, H. & Barabási, A.-L. *Nature* **406**, 378–382 (2000).
2. Erdős, P. & Rényi, A. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–60 (1960).
3. Huberman, B. A. & Adamic, L. A. *Nature* **401**, 131 (1999).
4. Albert, R., Jeong, H. & Barabási, A.-L. *Nature* **401**, 130–131 (1999).
5. Faloutsos, M., Faloutsos, P. & Faloutsos, C. *Comput. Commun. Rev.* **29**, 251–263 (1999).
6. Kauffman, S. A. *The Origins of Order* (Oxford Univ. Press, 1993).
7. Claffy, K., Monk, T. E. & McRobb, D.
<http://helix.nature.com/webmatters/tomog/tomog.html>

Cell cycle

Conducting the mitotic symphony

David Cortez and Stephen J. Elledge

The process by which chromosomes condense and segregate during mitosis (nuclear division) can be likened to a symphony in which many instruments, working individually, are coordinated to produce a collective piece of elegance and beauty. The conductor, with a wave of the baton, ensures that each musical instrument enters the symphony at the proper time. The conductors of the mitotic symphony, called 'checkpoints', do much the same thing, and prevent errors in chromosome segregation that can lead to diseases such as Down's syndrome and cancer. On page 430 of this issue¹, Scolnick and Halazonetis identify a new conductor that monitors early movements in the mitotic symphony, and describe a possible link between cancer and a failure of this conductor to do its job.

Before mitosis in an animal cell can

proceed, the chromosomes must duplicate and the cell must build a 'spindle' — a machine, based on cytoskeletal filaments called microtubules, that will transport the chromosomes into the two daughter cells (Fig. 1). The spindle forms outside the nucleus, between two microtubule-organizing centres (centrosomes). This dual origin of the spindle means that it is bipolar. As the spindle forms, the chromosomes within the nucleus begin to condense into a more compact form that allows them to move more easily; this phase of mitosis is called prophase. The nuclear 'envelope' then breaks down during prometaphase, allowing each pair of duplicated chromosomes to attach to the spindle. The chromosomes align in a plane between the spindle poles during metaphase, and await a signal that abruptly separates the

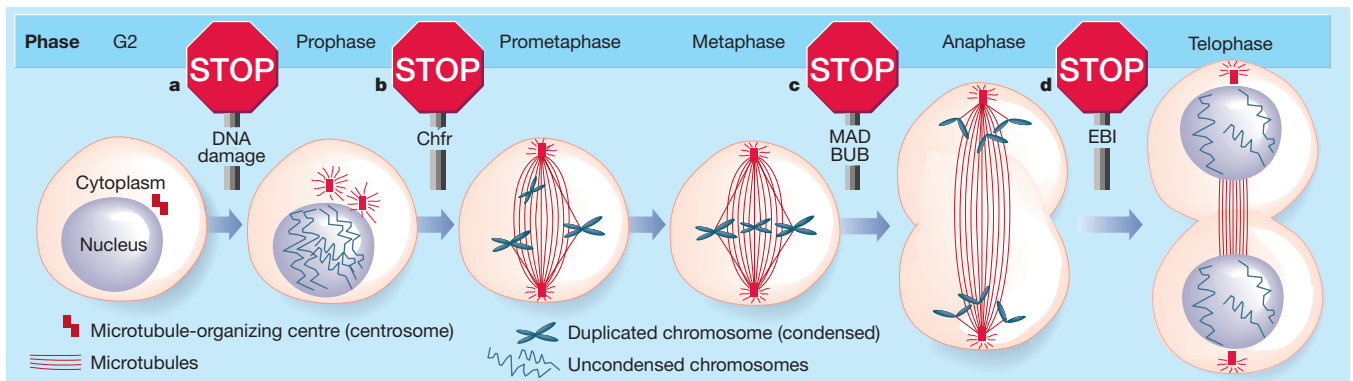


Figure 1 The conductors of the mitotic symphony. The phases of nuclear division (mitosis) in animal cells are shown. Mitosis is regulated by at least four conductors, called checkpoints. a, The DNA-damage checkpoint delays entry into mitosis until any damaged DNA is repaired. b, Scolnick and Halazonetis¹ have shown that, in response to microtubule poisons, a checkpoint dependent on the Chfr protein delays chromosome

condensation during prophase. c, The MAD/BUB-dependent checkpoint prevents chromosome separation and entry into anaphase until the chromosomes are attached correctly to the microtubule-based chromosome-segregation machinery (the spindle). d, At least in yeast, EBI participates in a checkpoint that delays the final production of two daughter cells in response to an incorrectly orientated spindle.

chromosomes of each pair, marking the start of anaphase.

We already knew about the existence of two conductors of this mitotic symphony (Fig. 1). One, the spindle-assembly checkpoint, ensures that each pair of chromosomes is correctly attached to a bipolar spindle before anaphase². A second conductor in yeast depends on the *BIM1/EB1* gene³; this checkpoint delays the exit from mitosis when the spindle is orientated abnormally.

Scolnick and Halazonetis¹ have now defined a third mitotic conductor, which coordinates an earlier part in the process. This conductor delays chromosome condensation in response to mitotic stress induced by microtubule poisons. The existence of this checkpoint was suggested by earlier experiments⁴ in which a delay in chromosome condensation in response to a microtubule poison was seen in normal human cell lines but not in several tumour cell lines. Scolnick and Halazonetis have now identified a gene — *chfr* — that seems to be required for delaying prophase in human cells. The sequence of the *chfr* gene is similar to that of the fission yeast *DMA1* gene, which is involved in a later mitotic checkpoint that delays a cell's exit from mitosis in response to spindle damage⁵.

Scolnick and Halazonetis also studied eight different human tumour cell lines, and found that the Chfr checkpoint did not function in four of them. Three of these cell lines do not express the *chfr* gene; in the fourth cell line, this gene contains a missense mutation, which results in the production of a defective copy of the Chfr protein. When the authors transfected the cell lines with normal or mutant *chfr*, they found that the wild-type — but not the mutant — gene restored the checkpoint, showing that *chfr* is essential for this checkpoint.

It is not yet known which microtubule-dependent process is coordinated by the Chfr checkpoint, but during prophase an

array of microtubules extends from the centrosomes, which then migrate to opposite poles of the cell (Fig. 1). Chfr may coordinate these events with chromosome condensation, although it is unclear why this should be necessary. Perhaps it is actually the breakdown of the nuclear envelope that needs to be coordinated with centrosome separation: if the envelope is dissolved before the centrosomes migrate, the chromosomes might associate prematurely with the centrosomes, interfering with subsequent steps in mitosis.

How does this new conductor delay entrance into mitosis? Entry into mitosis requires the activation of a complex containing a protein kinase (called cdc2) and its regulatory subunit (cyclin B). Cyclin B/cdc2 activity remains high during the Chfr-dependent delay¹, so its activity is not the target of Chfr. Perhaps Chfr instead controls the localization of cyclin B to the nucleus, which is required for mitosis. In the fungus *Aspergillus nidulans*, the kinase NIMA appears to be required for mitotic entry after cdc2 is activated. NIMA is also needed for nuclear localization of *A. nidulans* cyclin B/cdc2 (ref. 6). So the human NIMA protein is one plausible target of the Chfr checkpoint.

Another possibility is that Chfr is involved in controlling the degradation of a protein that regulates chromosome condensation and nuclear-envelope breakdown. Chfr contains a domain that is found in several proteins involved in tagging other proteins with the peptide ubiquitin⁷, a label that marks proteins to be degraded. Chfr also contains an FHA domain, which may bind phosphorylated peptides⁸, so protein kinases might act upstream of Chfr.

What happens to the cell when this mitotic conductor does not do its job properly? Cells that lack *chfr* are more sensitive to microtubule poisons and exhibit abnormal nuclear morphology after treatment with them — an effect that might result from a

failure of checkpoint function. In principle, the level of expression of *chfr* or its ability to function could be used as a diagnostic marker. Chemotherapeutic protocols that exploit this sensitivity to microtubule poisons could then be designed for use in cancer treatment. Indeed, several classes of approved anti-cancer drug work by disturbing microtubule function⁹.

Half of the human tumour cell lines studied by Scolnick and Halazonetis did not express wild-type Chfr protein, so this checkpoint might be inactivated during the development of some tumours — a suggestion also made about other mitotic checkpoint proteins^{10–13}. But it is not clear how inactivating Chfr might contribute to tumour formation. The Chfr-deficient cells retain the spindle-assembly checkpoint¹, and at least two of the cell lines used by Scolnick and Halazonetis exhibit low rates of chromosome loss¹⁴. (High rates of chromosome loss are characteristic of many tumours and may contribute to the inactivation of tumour-suppressor genes.) Although the absence of *chfr* expression in tumour lines might reflect the loss of the gene in those cells, it is also possible that *chfr* was never expressed in these types of tissue. So a firm connection between *chfr* and cancer remains to be established, but the discovery of this new mitotic conductor provides much needed insight into how the suite of mitotic checkpoints ensures the harmony of the mitotic minuet.

David Cortez and Stephen J. Elledge are in the Verna and Marrs McLean Department of Biochemistry and Molecular Biology, and the Howard Hughes Medical Institute, One Baylor Plaza, Baylor College of Medicine, Houston, Texas 77030, USA.

e-mail: selledge@bcm.tmc.edu

1. Scolnick, D. M. & Halazonetis, T. D. *Nature* **406**, 430–435 (2000).

2. Amon, A. *Curr. Opin. Genet. Dev.* **9**, 69–75 (1999).

3. Muhua, L., Adames, N. R., Murphy, M. D., Shields, C. R. & Cooper, J. A. *Nature* **393**, 487–491 (1998).

4. Jha, M. N., Bamberg, J. R. & Bedford, J. S. *Cancer Res.* **54**, 5011–5015 (1994).
5. Murone, M. & Simanis, V. *EMBO J.* **15**, 6605–6616 (1996).
6. Wu, L., Osmani, S. A. & Mirabito, P. M. *J. Cell Biol.* **141**, 1575–1587 (1998).
7. Freemont, P. S. *Curr. Biol.* **10**, R84–R87 (2000).
8. Durocher, D. *et al. Mol. Cell* **4**, 387–394 (1999).

9. Jordan, M. A. & Wilson, L. *Curr. Opin. Cell Biol.* **10**, 123–130 (1998).
10. Cahill, D. P. *et al. Nature* **392**, 300–303 (1998).
11. Jin, D. Y., Spencer, F. & Jeang, K. T. *Cell* **93**, 81–91 (1998).
12. Li, Y. & Benezra, R. *Science* **274**, 246–248 (1996).
13. Yamaguchi, K. *et al. Cancer Lett.* **139**, 183–187 (1999).
14. Lengauer, C., Kinzler, K. W. & Vogelstein, B. *Nature* **386**, 623–627 (1997).

Global change

Deciphering methane's fingerprint

Helmut Weissert

Two hundred years ago Georges Cuvier came up with the concept of what we now call 'proxy indicators' in geology. These are biological, chemical or physical signatures preserved in the rock record that serve as fingerprints of the biosphere's evolution through time. Much of the detective work on Earth's history has depended on these fingerprints.

These days, increasingly sophisticated tracer techniques are used in combination, and their robustness is tested by simulations of past global change. This is the approach that Hesselbo *et al.*¹ have chosen in the study described on page 392 of this issue — a carbon-isotope analysis of fossil wood deposited in near-shore sediments 180 million years ago during the Toarcian Stage of the Jurassic (200–140 million years ago). The authors have identified an atmospheric carbon-isotope signal in their samples that records an anomaly in the atmospheric carbon system; moreover, it is coupled to a known perturbation of the marine carbon reservoir. Hesselbo *et al.* propose that a massive release of methane, stored in gas hydrates beneath the sea floor, caused the perturbation in the carbon-isotope record, and that the methane pulse triggered major changes in climate and the oceans.

These conclusions are based mainly on carbon-isotope geochemistry. The relative concentrations of the two stable carbon isotopes (carbon-12 and carbon-13) in organic matter and biologically generated carbonate are used as fingerprints that reflect events in the carbon system. Marine sediments deposited over the past 600 million years retain the carbon-isotope signature from the time of their deposition, and oscillations in the geological record show that marine carbon-isotope composition was not stable². Differences in that composition can be related either to changes in the carbon isotopes entering the oceans, or to isotopic variations in total carbon sedimentation, which are caused mainly by varying rates of organic-carbon burial in sediments on the sea floor.

Organic carbon is enriched in the light isotope, carbon-12. So, at times of high rates of burial, carbon-13 became more abundant in the ocean water and is reflected as a positive

carbon-isotope anomaly in the rock record left by marine sediments. Burial of excess organic carbon was initiated by a warm, greenhouse climate, and coincided with crises in biocalcification in reefs and in calcareous plankton³, which were triggered by high levels of atmospheric CO₂ (ref. 4) and related changes in ocean chemistry and climate. Both processes contributed to a lowering of the amount of CO₂ in the atmosphere.

Negative carbon-isotope anomalies record episodes of oceanic carbon-12 enrichment in the past. Extraordinary carbon-12 enrichment of the marine carbon reservoir within only a few tens of thousands years coincided with known environmental catastrophes, as at the Cretaceous/Tertiary boundary⁵, some 65 million years ago. A collapse of the marine ecosystem and the con-

sequent interruption of the marine carbon flux into sediments may have caused the observed carbon-12 enrichment⁵. Other negative anomalies are evident at the base of positive carbon-isotope events⁶, and have been considered as the fingerprint of rapid influx of volcanic CO₂ enriched in carbon-12. This volcanic CO₂ would have triggered the biotic changes recorded in the subsequent positive carbon-isotope anomalies.

A few years ago, however, Dickens *et al.*⁷ proposed a new explanation for the negative carbon-isotope anomaly at the Palaeocene/Eocene boundary 55 million years ago. They calculated that only a sudden influx of methane derived from sources under the sea floor could account for the anomaly. Methane is enriched in the light isotope, carbon-12. It is stored in sedimentary gas hydrates beneath the oceans, and consists of solid crystals of water and methane that are stable under a wide range of temperatures and pressures.

Dickens *et al.* argued that deep-water warming or a rapid pressure decrease (or both) resulted in sudden release of methane (Fig. 1). The methane was oxidized to CO₂, which was enriched in methane-derived carbon-12. Higher CO₂ concentrations in oceans and the atmosphere were a consequence of this environmental catastrophe. Plants take up CO₂ in photosynthesis, so a pulse of methane into the atmosphere should be evident in the terrestrial carbon-isotope records of organic matter.

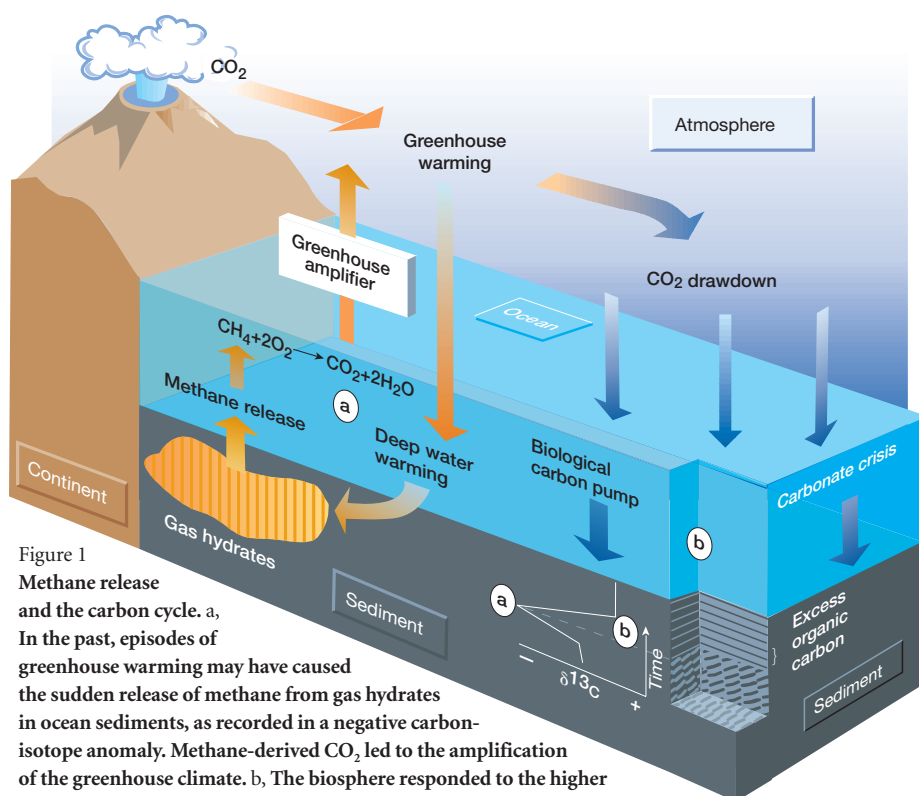


Figure 1 Methane release and the carbon cycle. **a**, In the past, episodes of greenhouse warming may have caused the sudden release of methane from gas hydrates in ocean sediments, as recorded in a negative carbon-isotope anomaly. Methane-derived CO₂ led to the amplification of the greenhouse climate. **b**, The biosphere responded to the higher CO₂ levels with accelerated burial of organic carbon on the ocean floor, and with crises in carbonate production, as recorded in positive carbon-isotope anomalies. Both processes contributed to a lowering of atmospheric CO₂ levels. Hesselbo *et al.*¹ have identified a terrestrial isotopic signal that provides evidence of a massive release of methane about 180 million years ago.

This is what Hesselbo and co-authors¹ have looked for, taking as their target one of the most prominent negative carbon-isotope anomalies known from the Jurassic. This anomaly occurs at the base of a large positive carbon-isotope 'excursion', which corresponds to the Toarcian oceanic anoxic event; this was a time when the deep-ocean waters were severely depleted in oxygen and there was excess burial of organic matter. The terrestrial carbon-isotope curve measured by Hesselbo *et al.* from their wood samples shows that the negative marine carbon-isotope anomaly is also reflected in the record of the global carbon cycle.

So what was the cause? The negative carbon-isotope event coincides with an episode of strong volcanic activity. But mass-balance calculations show that volcanic carbon cannot by itself account for the amount of carbon-12, which is calculated to have been generated within a few tens of thousands of years. Instead, Hesselbo *et al.* conclude that the Toarcian carbon-isotope event was a consequence of sudden methane release from beneath the ocean floor. Climate warming triggered by CO₂ from volcanoes could have led to a sudden warming of the deep ocean and the destabilization of gas hydrates. The authors think that a huge amount of methane was released, corre-

sponding to about 20% of the present-day gas-hydrate reservoir and twice that emitted at the end of the Palaeocene.

The work of Hesselbo *et al.*, and of others⁸, indicates that during the Mesozoic (250–65 million years ago) a warmer, greenhouse climate may have favoured episodic release of gas hydrates into the atmosphere. This raises the provocative question of whether the carbon-isotope anomalies at the Cretaceous/Tertiary boundary, and also at the 250-million-year-old Permian/Triassic boundary, were a result of sudden methane release — in these cases possibly triggered by an extraterrestrial object hitting the Earth. As the accuracy of our information from proxy indicators increases, Mesozoic environmental history is becoming an exciting arena for high-resolution detective work. ■

Helmut Weissert is in the Earth Science

Department, ETH, CH-8092 Zürich, Switzerland.
e-mail: helmi@erdw.ethz.ch

1. Hesselbo, S. P. *et al.* *Nature* **406**, 392–395 (2000).
2. Veizer, J. *et al.* *Chem. Geol.* **161**, 59–88 (1999).
3. Weissert, H. *et al.* *Paleogeog. Paleoclimatol. Paleoceanogr.* **137**, 189–203 (1998).
4. Larson, R. L. *Geology* **19**, 547–550 (1991).
5. Hsü, K. J. & McKenzie, J. A. *Am. Geophys. Union Monogr.* **32**, 487–492 (1985).
6. Menegatti, A. *et al.* *Paleoceanography* **13**, 530–545 (1998).
7. Dickens, G. R. *et al.* *Paleoceanography* **10**, 965–971 (1995).
8. Jarhen, A. H. & Arens, N. C. *Geol. Soc. Am. Abstr. Progs* **30**, 53 (1998).

Developmental biology

Fringe benefits to carbohydrates

Mark E. Fortini

The past decade has seen tremendous advances in our understanding of the molecular mechanisms underlying development. Papers on pages 369 and 411 of this issue^{1,2}, along with studies in *Nature Cell Biology*³ and *Current Biology*⁴, provide fascinating insights into the biochemical features of one such mechanism, which establishes spatial boundaries in developing tissues. The authors make a convincing case that carbohydrate chains attached to the extracellular portion of a particular receptor on the cell surface can modulate the interaction of the receptor with its binding partners. Remarkably, the carbohydrates thereby lead to the receptor being activated in only a particular tissue area.

The proteins producing these carbohydrate modifications are Fringe — found in the fruitfly *Drosophila melanogaster* — and its vertebrate counterparts, Lunatic Fringe, Radical Fringe and Manic Fringe. Notch is the receptor in question, and it is activated at the cell surface when it binds to a ligand from the Delta or the Serrate/Jagged protein families. *Notch* was one of the first genes to be identified in *Drosophila*, being

described by T. H. Morgan and colleagues as early as 1916. It has long been known to control tissue-patterning events in which certain cells are singled out from a group of equivalent cells⁵. More recently, Notch has captured the attention of biologists studying

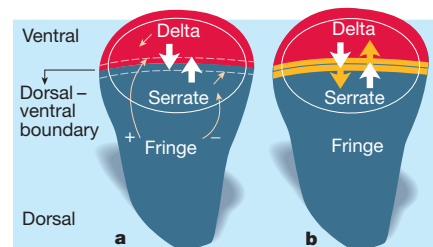


Figure 1 Creation of simple boundaries through complex carbohydrates. Early development of the *Drosophila* wing imaginal disc, an undifferentiated structure from which the wing will arise. a, The Notch protein and one of its ligands, Delta, are expressed widely in the disc. Serrate (another Notch ligand) and Fringe (a protein that modifies Notch's response to its ligands) are restricted to the dorsal compartment. Fringe inhibits the ability of Serrate to activate Notch (minus symbol),

vertebrate development, in part because malfunctioning of the Notch signalling pathway underlies an ever-growing list of human diseases, including T-cell leukaemia, adult-onset dementia, and the developmental disorders spondylocostal dysostosis and Alagille syndrome.

In certain situations during development, Notch signalling can be spatially modulated by Fringe, such that Notch is activated along a discrete line of cells. This specialized adaptation of the pathway is vital in establishing a sharp boundary between dorsal and ventral compartments during development of the insect wing, patterning of the chick limb bud, and other developmental processes. Genetic studies have led to an elegant model for how Fringe regulates boundary formation^{6–8} (Fig. 1). The key feature of this model is that Fringe somehow causes Notch to become more sensitive to Delta and less sensitive to Serrate/Jagged. As a result, Notch signalling is amplified in a narrow band of cells running along the dorsal–ventral boundary of the wing. But how does Fringe do this?

On the basis of slight similarities of the amino-acid sequence of Fringe to those of some bacterial enzymes⁹, Fringe has been proposed to be a glycosyltransferase — an enzyme that adds sugar residues to proteins. But, until now, direct evidence to support this suspicion was lacking.

Moloney *et al.*¹ and Brückner *et al.*² now show that Fringe is indeed a glycosyltransferase, with a new *N*-acetylglucosaminyl-transferase activity. Within the extracellular portion of Notch are 36 consecutive motifs called epidermal-growth-factor (EGF)-like domains. A five-carbon sugar, fucose, is linked to the side chains of serine, threonine or hydroxylysine amino acids within a subset of these domains (sugars linked to any of these three amino acids are known as *O*-linked sugars). Fringe causes elongation of these *O*-linked sugars to more complex forms.

Moloney *et al.*¹ transfected mammalian

whereas it potentiates the ability of Delta to activate Notch (plus symbol). Notch is thus activated most strongly in the area between the dashed lines: it is activated by Serrate in the Fringe-negative ventral cells along the boundary, and by Delta in the Fringe-positive dorsal cells along the other side of the boundary. b, As wing development proceeds, amplification of Notch signalling (yellow band) in cells on either side of the boundary results in a transcriptional feedback loop (yellow arrows) that eventually restricts ligand expression to the boundary (not shown). This system has similar functions in development in other organisms. New results^{1–4} show that the *Drosophila* and mammalian Fringe proteins modify the response of Notch to its ligands by adding sugar groups to Notch's epidermal-growth-factor (EGF)-like extracellular domains (see Fig. 6 on page 373 and Fig. 4 on page 413).

Notch1, together with Manic Fringe, into Chinese hamster ovary cells, and analysed the resulting spectrum of O-linked sugars (these rare glycosylation patterns had already been seen in a particular class of EGF motif found in Notch¹⁰). The authors found that Manic Fringe adds N-acetylglucosamine to fucose through an unusual linkage. They confirmed their results by *in vitro* assays using different Notch1 constructs, and showed that the effects of Fringe on Notch signalling depend on O-linked — but not N-linked (asparagine-linked) — glycosylation pathways.

Brückner *et al.*² took a different approach. They prepared cellular fractions containing recombinant Fringe, and showed that they specifically catalyse the addition of N-acetylglucosamine onto fucose, and not the transfer of other donor sugars onto other acceptor sugars. They also conclude that Fringe extends O-linked fucose chains on Notch. Supporting evidence comes from Munro and Freeman⁴, who show that Fringe — like many known glycosyltransferases — can be crosslinked to the nucleoside diphosphate UDP. And Fringe's active-site motif, DxD (where D denotes asparagine and x denotes any amino acid), is essential for Fringe's function in biochemical assays^{1,2} and transgenic flies^{1,4}.

Where does glycosylation of Notch by Fringe take place? Early studies indicated that Fringe might be secreted from the cell, although this idea was always difficult to reconcile with genetic data showing that Fringe acts on Notch in the same cell. But it seems that Fringe — like other glycosyltransferases — actually functions in the Golgi apparatus^{2–4}, an organelle in the cellular protein-secretory pathway. Fringe localizes to the Golgi in mammalian fibroblast cells³ and *Drosophila* cells^{2,4}. Forced secretion of Fringe out of the cell reduces its ability to function *in vivo*⁴, whereas its forced retention in the Golgi has few or no consequences². Hicks *et al.*³ show that Fringe must be expressed together with mammalian Notch1 in the same cell to have an effect on Notch signalling. It has no effect when expressed by the ligand-presenting cell or in a soluble secreted form. Brückner *et al.*² note similar effects of *Drosophila* Fringe. So, at least for full-length Notch, effective O-linked glycosylation might need to occur during its folding or maturation in the Golgi. This interaction between Notch and Fringe is presumed to be transient, contrasting with evidence that the two proteins associate in a stable complex¹¹.

Some details remain to be hammered out. For example, does Fringe affect the binding of ligands to Notch, or some other aspect of the ligand–receptor interaction? Brückner *et al.*² report that Fringe dramatically improves the binding of Delta to Notch, but were unable to detect the binding of Serrate, irrespective of the presence of Fringe. Moloney *et al.*¹

and Hicks *et al.*³ show, however, that Fringe has no measurable effect on ligand binding to Notch, and suggest that glycosylation may instead influence ligand-induced conformational changes in Notch. This discrepancy might reflect the fact that Brückner *et al.* worked mainly with Delta, while Moloney *et al.* and Hicks *et al.* concentrated mainly on Serrate/Jagged. Alternatively, it might result from differences in methodology.

This body of work has finally settled the question of whether or not Fringe has glycosyltransferase activity, and offers a stunning example of how an important signalling pathway can be modulated by differential receptor glycosylation. Although Fringe is clearly not essential for Notch signalling, the glycosylation of signalling components is likely to be of widespread relevance. Both *Drosophila* and mammals possess proteins with limited similarity to Fringe and with distinct glycosyltransferase activities or substrate preferences. The human disease multiple exostosis is caused by mutations in glycosyltransferases¹², and deficiencies in fucose metabolism are known to underlie leukocyte-adhesion deficiency type II¹³.

Perhaps we will also find mutations or variants of the human Fringe genes that result in human diseases. Such an event would surely lead to increased appreciation of the importance of protein–carbohydrate interactions in signal transduction and human biology. ■

Mark E. Fortini is in the Department of Genetics, University of Pennsylvania School of Medicine, Stellar-Chance Laboratories 709C, 422 Curie Boulevard, Philadelphia, Pennsylvania 19104, USA.

e-mail: fortini@mail.med.upenn.edu

1. Moloney, D. J. *et al.* *Nature* **406**, 369–375 (2000).
2. Brückner, K., Perez, L., Clausen, H. & Cohen, S. *Nature* **406**, 411–415 (2000).
3. Hicks, C. *et al.* *Nature Cell Biol.* **2**, 515–520 (2000).
4. Munro, S. & Freeman, M. *Curr. Biol.* **10**, 813–820 (2000).
5. Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. *Science* **284**, 770–776 (2000).
6. Panin, V. M., Papayannopoulos, V., Wilson, R. & Irvine, K. D. *Nature* **387**, 908–912 (1997).
7. Fleming, R. J., Gu, Y. & Hukreide, N. A. *Development* **124**, 2973–2981 (1997).
8. Klein, T. & Martinez-Arias, A. *Development* **125**, 2951–2962 (1998).
9. Yuan, Y. P., Schultz, J., Mlodzik, M. & Bork, P. *Cell* **88**, 9–11 (1997).
10. Moloney, D. J. *et al.* *J. Cell Biol.* **275**, 9604–9611 (2000).
11. Ju, B.-G. *et al.* *Nature* **405**, 191–195 (2000).
12. Selleck, S. B. *Trends Genet.* **16**, 206–212 (2000).
13. Karsan, A. *et al.* *J. Clin. Invest.* **101**, 2438–2445 (1998).

Global change

Silica control of carbon dioxide

Paul Tréguer and Philippe Pondaven

There is a popular view that the availability of iron, a trace nutrient, is a prime determinant of phytoplankton productivity in the oceans. But there are plenty of other influences on this primary productivity. Writing in *Paleoceanography*¹, Kevin Harrison looks at the role of silica and argues that variations in silica's oceanic cycle would have altered the species composition of phytoplankton. The implications are considerable: such alterations are likely to have had a large effect on levels of CO₂ in the atmosphere.

Dissolved CO₂ in the surface ocean is taken up by phytoplankton through photosynthesis, a process that requires nutrients such as nitrate, phosphate and silicate (dissolved silica), and trace metals that are available in surface waters only in limited amounts. Through respiration and microbial activity, CO₂ is then released back to the atmosphere. In principle, CO₂ uptake from and release to the atmosphere are balanced. But when more of the photosynthetic production is exported from the surface ocean to the depths, in the form of the remains of phytoplankton and other organic matter, more carbon is retained in the ocean and less returned to the atmosphere. This flux of carbon to the deep ocean is called the biological carbon pump.

It has already been hypothesized² that, in the equatorial Pacific, this pump is triggered by the availability of silicate. Harrison now takes the idea a lot further. He shows how variations in silicate inputs to the surface ocean can control levels of atmospheric CO₂ on a global scale. In particular, he considers that silicate controls can explain the rise in atmospheric CO₂ that occurred at the end of the Last Glacial Maximum, about 18,000 years ago, during the transition from glacial to interglacial conditions.

During the Last Glacial Maximum, atmospheric concentrations of CO₂ were 40% lower than they are now. Cores from polar ice sheets and from deep-sea sediments have revealed that more dust was delivered from the land to the ocean during glacial times. Dust contains iron, an essential trace metal for various marine microscopic algae. Hence the case for the 'iron hypothesis'³, which holds that variations in atmospheric CO₂ are related to iron inputs to the ocean surface. In principle, the more iron delivered to the ocean, the greater is the photosynthetic activity and export of carbon from surface to deep waters. In turn, this implies that more CO₂ is extracted from the atmosphere than is returned to it, the balance being locked up in carbon sequestered in the ocean depths.

The iron hypothesis has been useful in

	Modern ocean	Last Glacial Maximum
Wind-borne silica	0.5	3.7
Primary production		
Carbon	3,827	3,827 (0%)
Silica	268	396 (+48%)
Export flux		
Carbon	241	320 (+33%)
Silica	134	198 (+48%)
Contribution to biomass (%)		
Diatoms	54	79
Non-siliceous phytoplankton	46	21

Figure 1 The oceanic effects of higher silica inputs during the Last Glacial Maximum as compared with modern (interglacial) times. Units are teramol yr^{-1} ($10^{12} \text{ mol yr}^{-1}$). The model from which these results come is based on that of Tyrrell⁷, with inclusion of the ocean silica cycle⁶; river–ocean nutrient fluxes are assumed to be the same for both cases. In accordance with Harrison's new work¹, wind-borne silica input was set at 7.4 times higher in the Last Glacial Maximum. The argument is that conditions favouring siliceous diatoms (meaning fewer CO_2 -generating non-siliceous species), and the increased export of carbon from surface waters to the deep ocean, can explain the lower levels of atmospheric CO_2 during the Last Glacial Maximum.

explaining the low primary productivity of the nutrient-rich modern equatorial Pacific and Southern Ocean, caused (so the argument runs) by iron limitation. But its validity in periods in the past is questionable. For instance, after the decrease in dust delivered to the ocean at the end of the Last Glacial Maximum it took about 8,000 years for levels of atmospheric CO_2 to increase to those typical of interglacial periods. Because the residence time of iron in the ocean is only a few tens of years⁴, iron alone cannot account for the variations in atmospheric CO_2 during this glacial–interglacial transition⁵. Consideration of conditions in the terrestrial biosphere during the transition, or of the physics of the ocean, cannot help much in explaining these variations¹. So answers must be sought in changes in ocean biogeochemistry.

Harrison¹ concentrates on silicate, which has a residence time of about 15,000 years⁶ — enough time to allow for the slow rise in atmospheric CO_2 after the decrease in dust delivery at the end of the Last Glacial Maximum. Harrison challenges the iron hypothesis. He points out that material transported by wind from land to sea also contains silica, which dissolves at least partly in the surface ocean. Silicate is needed by diatoms to build up their siliceous cell walls, and increasing silicate availability in the ocean favours the growth of diatoms over non-siliceous species of phytoplankton. A significant part of these non-siliceous species consists of coccoliths, which have

shells of calcite (calcium carbonate). Counterintuitively, the production of calcite (by reaction between dissolved calcium and hydrogen carbonate) results in the production of dissolved CO_2 , which can escape from the surface ocean to the atmosphere.

So dominance of diatoms over coccoliths means decreased calcite production in sea water and lowering of the flux of CO_2 from the surface ocean to the atmosphere. The result is a net decrease in atmospheric CO_2 . To account for 40% less atmospheric CO_2 in the Last Glacial Maximum than in the ensuing interglacial, Harrison needs a dust-borne silica flux to the ocean that is higher by a factor of about seven. That figure is supported by the evidence from marine sediments.

During the Last Glacial Maximum, how did the biological carbon pump respond to the higher inputs of dust? To answer this question, we carried out a complementary calculation to Harrison's¹. We used a simple biogeochemical two-box model, derived from that of Tyrrell⁷, incorporating the ocean silica cycle and different key species of phytoplankton. Assuming, like Tyrrell⁷, that phosphorus is the ultimate limiting nutrient, we found that increasing silicate dust inputs by a factor of about seven does not change the total primary production. But in the model, production becomes dominated by diatoms rather than non-siliceous species (Fig. 1). The generation of biogenic silica rises by about 50% over that of the best estimate for the modern ocean⁸, and the export flux of organic carbon, mainly from siliceous organisms, rises by 33%. All in all, then, reduced CO_2 output from non-siliceous organisms and higher carbon sequestration in the ocean depths can account for the lower levels of atmospheric CO_2 during the Last Glacial Maximum. Likewise, changes in silicate

control can explain the glacial–interglacial rise in CO_2 .

There are other considerations, however, which have not been taken into account here. During the Last Glacial Maximum, not only was the wind-borne flux of silica dust different to that under interglacial conditions, but so too was the delivery of nutrients from rivers into the ocean. River inputs of silicate to the modern ocean are an order of magnitude higher than those from wind-borne dust⁶. So any reduction in river inputs would greatly affect calculations based on Harrison's scheme. However, Froelich *et al.*⁹ have shown that the flux of silicate from rivers to oceans was even higher during the Last Glacial Maximum than it is now; that would further favour the dominance of diatoms over other phytoplankton species.

We have a long way to go before we have a secure appreciation of the links between nutrients, phytoplankton productivity and CO_2 in the atmosphere. But, at the least, Harrison's 'silica hypothesis' is a fresh illustration of how central marine biogeochemistry is in understanding the long-term variation of atmospheric CO_2 and of Earth's climate. ■

Paul Tréguer and Philippe Pondaven are at the Université de Bretagne Occidentale, UMR CNRS 6539 Institut Universitaire Européen de la Mer, Technopôle Brest-Iroise, Place Copernic, 29280 Plouzané, France.

e-mail: paul.treguer@univ-brest.fr

1. Harrison, K. G. *Paleoceanography* **15**, 292–298 (2000).
2. Dugdale, R. C. & Wilkerson, F. P. *Nature* **391**, 270–273 (1998).
3. Martin, J. H. & Gordon, R. C. *Deep-Sea Res.* **35**, 177–196 (1988).
4. Toggweiler, J. R. *Nature* **400**, 511–512 (1999).
5. Broecker, W. S. & Henderson, G. M. *Paleoceanography* **13**, 352–364 (1998).
6. Tréguer, P. *et al.* *Science* **268**, 375–379 (1995).
7. Tyrrell, T. *Nature* **400**, 525–531 (1999).
8. Nelson, D. M. *et al.* *Glob. Biogeochem. Cycles* **9**, 359–372 (1995).
9. Froelich, P. N. *et al.* *Paleoceanography* **7**, 739–767 (1992).

Earth science

The extraterrestrial wedding ring

Richard J. Walker

A topic of hearty debate amongst geochemists is the origin of the so-called highly siderophile ('iron loving') elements in the Earth's mantle. This is a group of elements that includes two of our most highly prized commodities, gold and platinum. Put simply, there isn't much of them in the mantle, but there is a lot more than might be expected. In the paper on page 396 of this issue¹, Holzheid *et al.* lend weight to a theory that invokes an extraterrestrial explanation for this puzzle.

Apart from gold and platinum, the siderophile elements include iridium, osmium, palladium, rhenium, rhodium and ruthenium. They are characterized by their

extreme affinity for metal (iron) magmas, and have a 10,000 times greater preference to partition into these magmas than the silicate magmas from which the mantle is largely formed. Consequently, during primary differentiation of the rocky planets into cores and mantles, these elements were almost totally concentrated in the metallic cores. In the Earth, for instance, about 99.99% of the planet's budget of highly siderophile elements resides in the core (which is not, alas, accessible for mining).

What Holzheid and colleagues have done is provide new experimental data suggesting that the highly siderophile elements were added to the Earth from meteorites after the

Earth had initially formed and the core had become distinct from the mantle. So, according to these results, the gold in your jewellery can be viewed as a relative latecomer to our planet.

The debate about the highly siderophile elements has centred on why they are as abundant in the mantle as they are. The evidence for thinking that their abundance should be much lower comes from chondritic meteorites. These are bits and pieces left over from the formation of the early Solar System that have fallen to Earth, and are the most primitive materials available for comparison. Chondritic meteorites have abundances of highly siderophile elements of between 150 and 300 times that of the mantle, yet the mantle concentrations predicted by metal–silicate magma partitioning experiments conducted at 1 atmosphere of pressure are more than two orders of magnitude lower². That the concentrations are not as low as predicted is fortuitous for jewellers (and others), because viable economic deposits of these elements would probably not have formed from a mantle with the much lower concentrations. One other curious aspect about the mantle abundances of the highly siderophile elements is that, on a coarse scale at least, they occur in ‘chondritic relative abundances’ — meaning that the ratio of the abundance of one element to another in the mantle is similar to the ratio in chondritic meteorites³.

Over the past 40 years, many models have been proposed to explain the abundances of these elements in the mantle. But recent discussion has largely settled on two possibilities. In one, it has been proposed that metal–silicate magma segregation may have occurred during the first 100 million years of the Earth’s history at the base of a deep magma ocean (extending to 500–1,000 km beneath the surface), and that the affinity of these elements for iron may decrease as a function of pressure, thereby better matching the observed concentrations^{4–6}.

A flurry of high-pressure partitioning experiments during the past decade has shown that this model is robust for at least explaining the abundances of some less-siderophile elements, such as nickel and cobalt⁵. The main constraint to this model is that a condition has to be found under which all of the highly siderophile elements have a low enough affinity for metallic iron to account for their abundance in the mantle. Those affinities must be nearly the same to account for the chondritic relative abundances. One study⁶ has reported that rhodium exhibits a greatly reduced affinity for iron at higher pressures, lending credence to this model. Until the work of Holzheid and colleagues¹, however, there have been few data for other highly siderophile elements.

The second possibility has been termed the ‘late veneer’ or ‘heterogeneous accretion’ model. This attempts to explain the mantle

abundances of the highly siderophile elements by appealing to the addition of 0.5–1% of oxidized chondritic material to the Earth after the core had formed^{7,8}. The relatively high abundances of these elements in chondrites would dramatically raise their concentrations in a primitive mantle that had effectively been stripped of them by core formation. The oxidized nature of the material would not allow the highly siderophile elements to move back into the core.

Holzheid *et al.*¹ describe results for the partitioning of platinum and palladium between metal and silicate magmas at high pressures (equivalent to depths of 500 km) and temperatures. Such experiments are notoriously difficult. In part this is because of the low concentrations of the elements present in silicate magma equilibrated with metal magma. More troublesome, however, is the tendency for tiny nuggets of the elements to form and not segregate readily from the silicate into the metal magma. This can lead to seemingly higher concentrations of the element in the silicate than is consistent with equilibrium partitioning.

For the latest experiments¹, palladium was chosen because of its limited tendency to form nuggets. Nugget formation is still an issue with platinum, but the study at least provides minimum constraints on its affinity for metallic iron. The results show that both palladium and platinum retain very high affinities for metal relative to silicate magma under high pressures and temperatures, and that, at least for the conditions tested, the process could not lead to the abundances of these two elements found in the mantle.

So the results cast doubt on the ability of the high-pressure metal–silicate segregation model to account for the mantle abundances of highly siderophile elements, and are consistent with the late-veneer model. They do not provide evidence against the purported deep segregation of metal and silicate magmas, however. Ultimately it may transpire that some elements in the mantle, such as the less-siderophile nickel and cobalt, owe their abundances to metal segregation at the base of a deep magma ocean, and that only the highly siderophile elements are dominated by the late-veneer signature⁹. ■

Richard J. Walker is in the Department of Geology, University of Maryland, College Park, Maryland 20742, USA.

e-mail: rjwalker@geol.umd.edu

- Holzheid, A., Sylvester, P., O'Neill, H. St C., Rubie, D. C. & Palme, H. *Nature* **406**, 396–399 (2000).
- Jones, J. H. & Drake, M. J. *Nature* **322**, 221–228 (1986).
- Chou, C.-L., Shaw, D. M. & Crockett, J. H. *J. Geophys. Res.* **88**, A507–A518 (1983).
- Murthy, R. *Science* **253**, 303–306 (1991).
- Li, J. & Agee, C. B. *Nature* **381**, 686–689 (1996).
- Righter, K. & Drake, M. J. *Earth Planet. Sci. Lett.* **146**, 541–553 (1997).
- Chou, C.-L. *Proc. Ninth Lunar Planet. Sci. Conf.* 219–230 (Pergamon, New York, 1978).
- Morgan, J. W. *J. Geophys. Res.* **91**, 12375–12387 (1986).
- Drake, M. J. *Geochim. Cosmochim. Acta* **64**, 2363–2370 (2000).

Daedalus

Hidden thoughts

Last week Daedalus proposed a method of measuring brain activity, in neural pulses per second. Each pulse puts the sodium ions of the relevant nerve into brief 1-kilohertz oscillation. At that moment, their nuclear magnetic excitation at 1 kHz is strongly damped. DREADCO volunteers are now testing a nuclear magnetic psychometric hat, which measures the wearer’s changing sodium-ion relaxation density, locates the sources of this activity by magnetic imaging, and transmits the results to a base station.

The steady cerebral base load of bodily ‘housekeeping’, keeping the heart and lungs pumping, and so on, should be fairly constant. Processing due to perception and conscious thought will vary wildly, but in keeping with the subject’s circumstances. Sensory processing will be identified by its decline during sensory deprivation, and conscious thought by standard tasks and tests. The most interesting study will be subconscious thought. Even psychologists find it an elusive concept.

Daedalus feels that the subconscious mind cares little for the conscious needs of the moment. Thus a creative scientist can have a new idea at any time. The subconscious may have been developing the idea for ages before suddenly ‘pushing it upstairs’ for conscious consideration. With luck, one of Daedalus’s volunteers will have such an inspiration while wearing the psychometric hat, thus capturing the whole sequence. It should show the subconscious activity, the burst of inspiration, and the transfer of activity to a different region as the conscious mind considers the new idea.

The subconscious also handles personal problems. Volunteers wrestling with superego–id conflicts, or shouldering or rejecting guilt or anger, or arguing with internalized parent figures, should reveal the fact by extra heavy processing in specific regions.

Once identified, subconscious activity could be tamed and trained by biofeedback methods. Shown their brain activity in real time, the volunteers could learn how to control it. Chronic worriers will soothe away their fears; those sabotaged by repressed memories or fatigued by obsessive notions will learn, if not to damp them out, at least to run them efficiently with the minimum number of neurons. Even conscious thought could be neurologically streamlined. The mental space thus opened will give a new freedom and spaciousness to mental life, and make room for more creative imagination.

David Jones

Stem cells — hype and hope

Ron McKay

Studies of stem cells will help in understanding the development and function of organs in mammals. They may also offer a way of treating diseases ranging from liver failure to Parkinson's disease.

Politicians, journalists, patients... it seems as though everyone is talking about stem cells. Why all the fuss? Stem cells have become the centre of so much attention because they turn into all the different cell types that make up complex organisms, and they promise to perform this remarkable feat on demand. The implications for medicine are profound, but practical and ethical barriers stand in the way. Are our hopes for this technology justified?

Germ cells and cancer

All cells come from other cells, so what does it mean to be a stem cell? There are complex academic definitions, but the most important idea is that stem cells are undifferentiated — unspecialized — cells that can renew themselves and also give rise to one or more specialized cell types with specific functions in the body. This concept applies to many situations in biology and medicine. For example, tumours start off as just a single cell and, in many tumours, several cell types are derived from the tumour 'stem' cell. An extreme example of this phenomenon is seen in tumours of the testis called teratomas or teratocarcinomas. In mice, these tumours can be generated simply by taking normal testis cells and placing them at a new site. The observation that single tumour cells can generate many different cell types has had dramatic consequences.

Stem cells are defined by their developmental potential. The idea of a stem cell has a natural meaning in the context of the germ line; after all, the germ line of animals produces eggs or sperm, which together generate whole organisms. The fertilized egg, or zygote, is a 'totipotent' (from the Latin *totus*, meaning 'entire') stem cell, and generates all the cells of an animal — including those that do not form part of the embryo, such as the cells of the placenta. As development proceeds, cells become channelled into particular pathways of differentiation, and their developmental potential becomes modified. Stem cells in these pathways can differentiate easily into a limited number of cell types; for example, stem cells in the brain ultimately give rise to all the different types of neuronal and non-neuronal cells in the central nervous system.

It would seem reasonable to expect the existence of stem cells during development, but they also occur in adult tissues. For

example, adult muscle stem cells can rebuild a muscle, and haematopoietic stem cells in adults can restore all the different cell types found in blood.

Embryonic stem cells

Inspired by the work on teratomas, scientists soon realized that there is a restricted period during early mouse development when certain normal cells have a remarkable ability to differentiate into a huge variety of cell types. These early embryonic stem (ES) cells can be taken from the embryo and grown in the laboratory; when placed back into a developing embryo, they contribute to all of the tissues of the mouse — including the germ line¹. Such cells are said to be 'pluripotent' (*plures*, meaning several). The difference between the pluripotent ES cell and the totipotent zygote is that the ES cell can only generate cells that form part of the embryo itself. Another type of pluripotent cell is the embryonic germ (EG) cell. These come from a later stage in development, when the cells of the germ line become set aside.

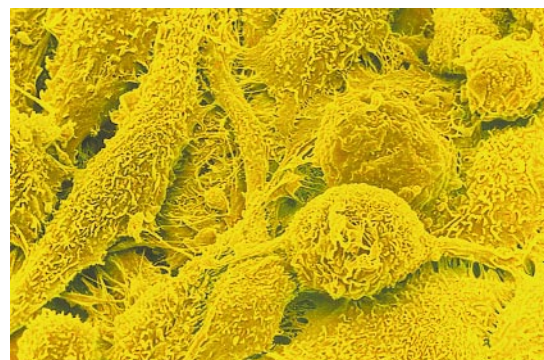
The ES and EG cells also exist in humans, and it has become possible over the past few years to grow them in the lab²⁻⁴. The mouse

ES cell can be easily defined by its ability to contribute to all the tissues of a developing embryo, as normal mice can be derived from ES cells. But it is illegal worldwide to derive a person from human ES cells, so the experimental definition of such a cell must be couched in different terms. Human ES cells are generally obtained from the extra embryos that are generated in fertility clinics. At a very early stage, when the embryo consists of only about a thousand cells, they can be separated from each other and grown in tissue culture. They can give rise to many different cell types in culture, so this is how they are defined — as pluripotent cells in tissue culture.

There is clear academic interest in culturing mouse ES cells and placing them back in developing embryos (Box 1). But what is the motivation for working with early human cells? Broadly speaking, there are two answers. First, these cells can be used to investigate features that are specific to early human development. Second, ES cells generate somatic cell types — the variety of non-reproductive cells that make up the human body. By studying ES cells, we may gain a deeper understanding of the process

Box 1 Learning about gene function

Mouse embryonic stem (ES) cells, shown here, can be grown in large numbers for long periods in the laboratory, and their genes can be altered in precise ways while they are being cultured. The effects of these genetic changes can then be assessed by making the cells differentiate into other cells, either in a culture dish or after placing the cells into the blastocyst — an early stage of development in mammals. The ES cells become incorporated into the blastocyst and then become a part of the developing mouse. The importance of this technique cannot be exaggerated. Defining the functions of the thousands of mammalian genes is a central goal of modern biomedical research, and mouse ES cells will help us to achieve it.



Precise genetic manipulation is not restricted to ES cells. In a study published last month²³, genes in sheep fibroblast cells were altered by genetic recombination. The nuclei of the fibroblast cells were then transferred into eggs from which the nuclei had been removed, and

the eggs were allowed to develop into sheep containing the desired genetic change. The benefits of genetically modifying farm animals in this way might include the ability to generate strains of cattle free of the protein agent underlying bovine spongiform encephalopathy (BSE). **R.M.**

YORGOS NIKAS/SPL

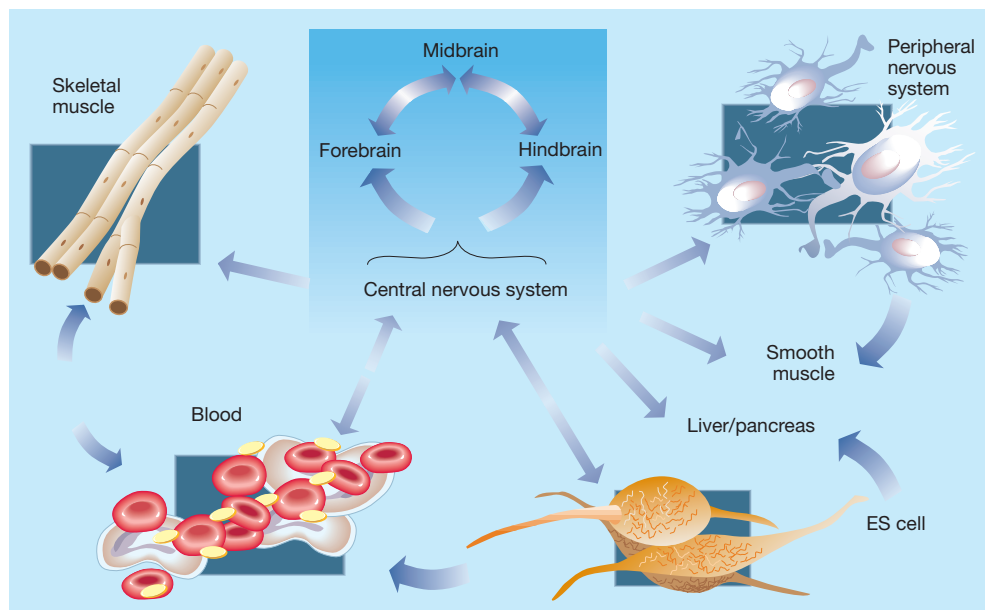


Figure 1 Stem-cell transitions. At least in the lab, stem cells are not always restricted to one particular pathway of differentiation. For example, central nervous system (CNS) stem cells form the different cell types of the CNS, but can also differentiate into haematopoietic (blood) stem cells. Blood stem cells in turn form the different cell types found in blood, as shown here, but can also differentiate into skeletal muscle stem cells (which differentiate into skeletal muscle cells, pictured) and central nervous system cells. Embryonic stem (ES) cells are pluripotent, and contribute to all of the tissues of developing mammals. For simplicity, only a few of the stem-cell types that ES cells can produce are shown here.

of cell replacement. The potential benefits to human health are huge, and range from generating new neurons for treating patients with Parkinson's disease to learning about the molecular processes that drive the development of tumours.

But the origins of human ES cells pose obvious ethical obstacles to their use in research (Box 2). One way round this problem might be provided by the observation that, during adulthood, the cells in most tissues are replaced — so there may be a source of new cells in the adult.

Adult stem cells

From the time that Henri Becquerel and Marie Curie discovered radiation, it has become increasingly clear that radiation damages rapidly dividing cells. It causes disease most readily in the white and red cells of

the blood (the haematopoietic system). These observations led eventually to the identification of the haematopoietic stem cell (HSC). An experimental definition of the HSC is that, amazingly, just one cell can reconstitute the entire blood system of a mouse⁵.

The fact that adult stem cells exist is exciting enough, but even more intriguing is that the potential of stem cells does not seem to be restricted by their source (Fig. 1). For example, a series of startling observations indicates that muscle and blood might be obtained from stem cells found in the tissues of either system^{6,7}. With all the genetic markers and protocols at our disposal for identifying HSCs, it will not take long to test this proposition — that exactly the same stem cells can generate the entire blood system as well as the striated muscle that allows you to run for the bus.

Stem cells are abundant in the developing brain and in two areas of the adult central nervous system (CNS): the hippocampus and the olfactory bulb. Cells in the brain are not replaced as efficiently as blood cells, however, so we cannot define adult CNS stem cells in the same way as haematopoietic stem cells. But CNS stem cells can be easily grown in the lab and — under the right conditions — differentiate efficiently in culture dishes into neurons, oligodendrocytes (the cells that insulate the electrical signals passing down axons in the nervous system) and astrocytes (another type of non-neuronal cell in the CNS). The many cell types of the peripheral nervous system are also generated from a stem cell.

Mouse CNS stem cells can also differentiate to form the cells of other organs, including muscle, blood, intestine, liver and heart^{8,9}.

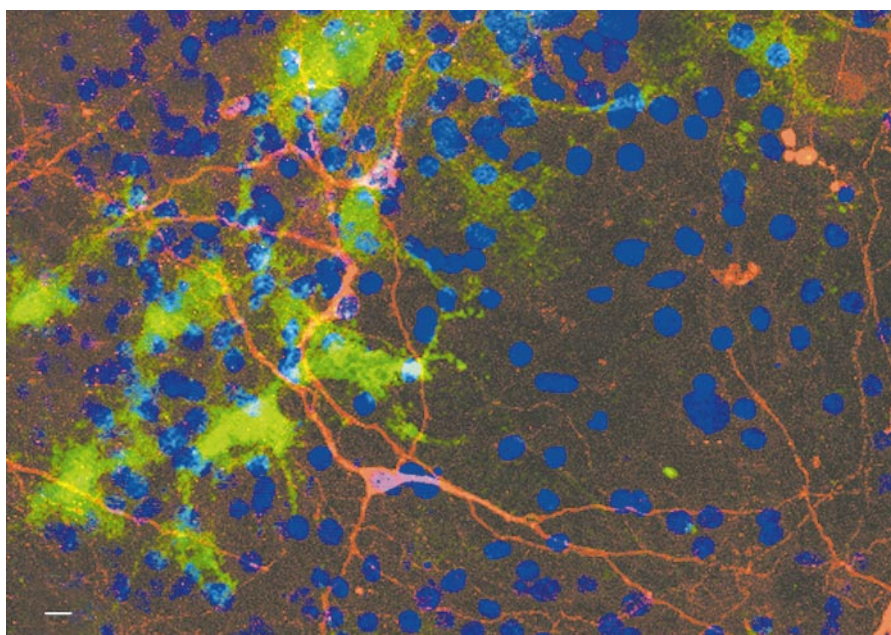


Figure 2 Dopamine-producing neurons derived from mouse embryonic stem (ES) cells¹⁵. ES cells (labelled with green fluorescent protein) were induced to differentiate into midbrain neurons that make the enzyme tyrosine hydroxylase (labelled with a red dye). These neurons make dopamine and form synapses with their normal brain targets when cultured together with those targets. Nuclei are labelled with a blue dye. The ability to generate dopamine-producing neurons from ES cells may provide an unlimited source of these neurons for clinical work and drug discovery. Scale bar, 20 μ m.

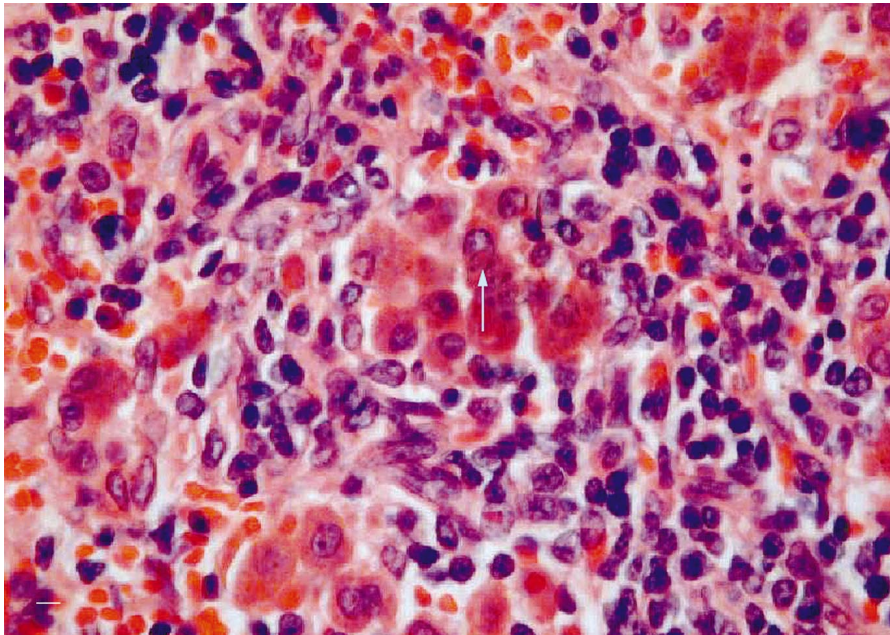


Figure 3 In principle, liver diseases might be treatable by transplanting isolated liver cells (hepatocytes). Donor livers, however, provide only a fraction of the cells needed for clinical transplantation, and primary liver cells do not proliferate significantly in tissue culture. When hepatocyte stem cells have been immortalized by introducing growth-promoting genes into them, they proliferate in culture, retain their ability to differentiate into hepatocytes, and function in animal models of liver failure¹⁶. This photomicrograph of a spleen section shows a transplanted cell (arrow) with hepatocellular morphology three months after transplantation. Scale bar, ~10 μm .

Conversely, blood cells can differentiate into brain cells¹⁰. It seems that adult stem cells may have the same developmental potential as ES cells, if given the right cues. But we know little about how stem cells can differentiate across boundaries, and how we could divert them into the pathway of choice. And — so far — no adult human stem cells have been shown to have such potential.

Mammals appear to contain some 20 major types of somatic stem cell. For example, stem cells have been described that can generate liver, pancreas, bone and cartilage. So it may be possible to obtain a wide range of stem cells from sources other than ES cells. This solution would effectively sidestep the ethical problem of using ES cells. But we have much to learn about how stem cells replicate and differentiate, and ES cells may offer essential technical advantages over other sources of somatic precursors.

Clinical potential and pitfalls

One reason for studying stem cells is their usefulness in cell-replacement therapy. The manipulation of HSCs is already an important clinical tool, and human HSCs have been essential in bone-marrow grafts that are in wide clinical use, for example in treating leukaemia patients. Although we are not certain of the identity of the stem cell concerned, skin cells can also be grown in large numbers, providing a life-saving grafting treatment for burn victims. And diabetes might be treatable by generating from stem cells the insulin-producing cells found in the normal pancreas and grafting them into the pancreas of diabetic patients.

Cell-replacement therapy also has potential in treating brain disease¹¹. When grafted into the developing or adult mouse brain, CNS stem cells can differentiate into neurons and glial cells (non-neuronal cells in the ner-

vous system) that are incorporated into the host tissues. Such grafted stem cells might be able to correct brain diseases characterized by loss of neurons. This idea is best developed for Parkinson's disease: there are encouraging findings showing that stem cells from the fetal brain could be used to restore some brain function to Parkinson's patients¹².

But we need routine access to the right cells to make cell grafting a practical technology. For example, the neurons required to treat Parkinson's disease can be obtained from the fetal brain, but there are technical and ethical barriers to using this tissue source. We can now generate the relevant neurons — those that produce the neurotransmitter dopamine — from mouse CNS stem cells, which can multiply and differentiate in the lab¹³. But the CNS stem cells quickly lose their ability to differentiate into dopamine-producing neurons. For this and many other cell types, it may not be possible to grow many stem cells in large enough numbers for cell-replacement strategies.

The pluripotent ES cell has an important advantage over somatic stem cells: it can be grown indefinitely in tissue culture. So it may be easier to obtain some of the cells that we need (such as dopamine-producing neurons for treatment of Parkinson's disease; Fig. 2) by culturing ES cells and prompting them to differentiate into the correct cell type. As yet, little is known about human ES cells, but many different cell types can be obtained from mouse ES cells in culture. Given the right combination of cues such as growth factors, these ES cells can differentiate *in vitro* into oligodendrocytes¹⁴ — the cells that are missing in patients with multiple sclerosis — and into the neurons that die during Parkinson's disease¹⁵. In humans,

oligodendrocytes can probably be obtained from sources less controversial than ES cells. But at the moment the only source of unlimited numbers of the neurons appropriate for Parkinson's patients is ES cells.

There may be other ways of obtaining enough adult stem cells for treatment purposes. While studying cancer, researchers have identified many genes involved in controlling cell growth. 'Immortalized' cells grow indefinitely in culture, without forming tumours when injected into animals. Many stem-cell types can be immortalized by introducing growth-promoting genes into them; after immortalization, they retain the ability to differentiate. Strikingly, immortalized cells can incorporate into the host tissue when they are grafted into animals. For example, earlier this year researchers showed that immortalized human liver cells, grafted into rats with acute liver failure (Fig. 3), can keep such rats alive¹⁶. There is a risk with using immortalized cells — they may be more likely to develop into tumour cells. A way around this problem is exemplified by the work with immortalized liver cells, which were generated with a 'cell-suicide' gene that could be activated by administering a drug to the rats. Strategies of this kind could allow cell numbers to be controlled and immortalized cells to become a clinically useful technology.

All in all, cell replacement is taking a central place in medicine, and the work discussed here is just a beginning. We do not know where unexpected benefits may suddenly emerge. For example, individual stem cells move easily through tissue and might be used to track down and kill cancer cells¹⁷. Another outcome of research into stem cells might be an understanding of how to direct the stem cells already present in our bodies to the necessary cell fates, without the need to

Box 2 Knowledge, ethics and laws

François Raspail understood as early as 1825 that cells are always derived from other cells ("Omnis cellula e cellula"; see ref. 24), that cells are not created equal, and that the mechanisms that generate differences between cells would explain the organization of cellular order in living organisms. These ideas provide the basis for modern studies of stem-cell biology. Knowledge gives us power, and the decision to act, or not to act, on the basis of that knowledge²⁵. A greater understanding of stem cells brings the power to treat diseases and we must decide how best to go about this.

Doubts arise mainly from the fact that many useful stem cells are derived from embryos or fetuses. Although in most cases the embryos would otherwise be discarded, the question remains as to whether the extraction of embryonic stem (ES) cells from such

embryos is justified. On the other side of the coin, this technology has great potential for the treatment of disease. There are alternatives: adult mouse stem cells (for example, muscle stem cells) can be diverted into different cell lineages. But their developmental potential might not be as great as that of ES cells, and it is difficult to extract and culture them.

Legislation around the world at present reflects this quandary. For example:

- In Germany, laws on reproductive medicine ban the extraction of stem cells from a human embryo.
- In the United States, federal funds cannot be used to conduct research using human pluripotent stem cells obtained from human fetal tissue or human embryos. The main federal funding agency, the National Institutes of Health, is investigating its guidelines on this issue at present. A bill

recently discussed by a Senate subcommittee proposed legislation that would allow researchers to obtain human ES cell lines with federal funding. At the time of writing, this bill had not yet been voted on. Private funding of research using ES cells is not prohibited in the United States.

- In Britain, it is illegal to derive ES cells from embryos or fetuses. But it is legal for UK researchers to import ES cells from abroad. The government is still debating whether or not to allow researchers to derive ES cells in the United Kingdom.
- French bioethics laws do not allow research on human embryos, but a recent report, adopted by the assembly general of the Conseil d'Etat, recommended that embryo research should be allowed for the purposes of research into stem cells. This issue is to be debated later this year. **R.M.**

and maintain tissues. We may also learn how to guide stem cells along desired pathways of differentiation, both *in vitro* and *in vivo*.

Conclusion

The wide distribution of stem cells raises some fundamental questions. Why are animals built this way? What other models of multicellularity can we imagine? Are there similar stem cells in model experimental organisms such as the fruitfly *Drosophila melanogaster* and the nematode worm *Caenorhabditis elegans*? These questions are interesting in their own right, and stem-cell biology will lead to a better understanding of the way in which animals work. But the history of modern biology is that even apparently obscure scientific work can bring practical benefit. Perhaps the answers to these fundamental questions will bring progress in transforming advances in stem-cell biology into a reduction in the burden of disease. When it comes to using human stem cells, particularly ES cells, there are clear ethical dilemmas (Box 2). But it is my belief that there is also a moral imperative in moving discoveries in this field to the clinic. ■

Ron McKay is at the National Institutes of Health, LMB/NINDS, Bethesda, Maryland 20892, USA.
e-mail: mckay@codon.nih.gov

1. Brook, F. A. & Gardner, R. L. *Proc. Natl Acad. Sci. USA* **94**, 5709–5712 (1997).
2. Thomson, J. A. *et al. Science* **282**, 1145–1147 (1998).
3. Shambloot, M. J. *et al. Proc. Natl Acad. Sci. USA* **95**, 13726–13731 (1998).
4. Reubinoff, B. E., Pera, M. F., Fong, C.-Y., Trounson, A. & Bongso, A. *Nature Biotechnol.* **18**, 399–404 (2000).
5. Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. *Science* **273**, 242–245 (1996).
6. Jackson, K. A., Mi, T. & Goodell, M. A. *Proc. Natl Acad. Sci. USA* **96**, 14482–14486 (1999).
7. Gussoni, E. *et al. Nature* **401**, 390–394 (1999).
8. Bjornson, C., Rietze, R., Reynolds, B., Magli, M. & Vescovi, A. *Science* **283**, 534–537 (1999).
9. Clarke, D. L. *Science* **288**, 1660–1663 (2000).
10. Eglitis, M. A. & Mezey, E. *Proc. Natl Acad. Sci. USA* **94**, 4080–4085 (1997).
11. Snyder, E., Taylor, R. & Wolfe, J. *Nature* **374**, 367–370 (1995).
12. Olanow, C. W., Kordower, J. H. & Freeman, T. B. *Trends Neurosci.* **19**, 102–109 (1996).
13. Studer, L., Tabar, V. & McKay, R. D. *Nature Neurosci.* **1**, 290–295 (1998).
14. Brüstle, O. *et al. Science* **285**, 754–756 (1999).
15. Lee, S.-H., Lumelsky, N., Studer, L., Auerbach, J. M. & McKay, R. D. *Nature Biotechnol.* **18**, 675–679 (2000).
16. Kobayashi, N. *et al. Science* **287**, 1258–1262 (2000).
17. Benedetti, S. *et al. Nature Med.* **6**, 447–450 (2000).
18. Cameron, H. A. & McKay, R. D. *Nature Neurosci.* **2**, 894–897 (1999).
19. Magavi, S. S., Leavitt, B. R. & Macklis, J. D. *Nature* **405**, 951–955 (2000).
20. Anderson, D. J. *Trends Genet.* **13**, 276–280 (1997).
21. McKay, R. *Science* **276**, 66–71 (1997).
22. Cory, S. *Nature* **401**, 538–539 (1999).
23. McCreath, K. J. *et al. Nature* **405**, 1066–1069 (2000).
24. Harris, H. *The Birth of the Cell* (Yale Univ. Press, New Haven, CT, 1999).
25. Monod, J. *Chance and Necessity* (Knopf, New York, 1971).

Related websites

- ♦ <http://www.aaas.org/spp/dspp/sfrl/projects/stem/main.htm>
- ♦ <http://bioethics.gov/pubs.html#stemcell>
- ♦ <http://www.nih.gov/news/stemcell/index.htm>
- ♦ <http://www.nuffield.org/bioethics/>

isolate and culture them. Such manipulation of resident stem cells may be clinically and ethically better than techniques in which cells are grown in the lab and grafted into the body where needed.

This may not be science fiction. In the brain, most neurons are generated during development, but, in the olfactory bulb and the hippocampus, neurons are still formed in the adult. The hippocampus is important in forming new memories, and, given recent success in restoring cell proliferation in this brain region in aged rats¹⁸, it might be possible to develop pharmacological methods to regulate the formation of new neurons in the hippocampus and to minimize memory loss during ageing. We might be able to apply this approach to many neurological problems, as replacement of neurons in the brain by the differentiation of intrinsic stem cells may be much more widespread than we previously thought¹⁹. But before we can recruit these resident stem cells, we need to know much more about the mechanisms that control their birth, fate and death.

Controlling cell fate

We are used to thinking of animal design as a consequence of the actions of an inexorable developmental machine. First there is the totipotent egg and then, through a series of irreversible restrictions in developmental

potential, the differentiated cell types of the animal are obtained. From this perspective, cells in adult tissues are seen as having few options. But, as discussed above, even adult animals contain stem cells with extraordinary developmental plasticity. This implies that the key events of development still occur in the 70-year-old heart or the 95-year-old brain.

The behaviour of stem cells depends on their history and on their local context or niche. They are not inflexible; we must simply learn how to influence them. Single extracellular signals that interact with cell-surface signal-detecting proteins can direct brain and blood stem cells to specific fates^{20–22}. For example, immature B cells (white blood cells) can be propagated for long periods in the presence of a cytokine called interleukin-7 (ref. 22). In the absence of interleukin-7 they differentiate into mature B cells. If a key gene transcription factor, Pax5, is missing from the immature B cells, they can be nudged into a variety of different cell types, such as dendritic cells (antigen-presenting cells) or macrophages (which engulf foreign material), depending on the combination of cytokines added to the culture medium. Isolating the fewer than 20 somatic stem-cell types, and defining their responses to external signals, will help us to decipher the language used to build

Five plus two equals yellow

Mental arithmetic in people with synaesthesia is not coloured by visual experience.

In synaesthesia, ordinary stimuli elicit extraordinary experiences. For example when C., who is a digit–colour synaesthete, views black digits, each number elicits a photism — a visual experience of a specific colour. It has been proposed that synaesthetic experiences differ from imagery in their consistency¹, automaticity^{2,3} and reliance on external stimuli to induce them⁴. Here we demonstrate that C.'s photisms are both consistent and automatic, but we find that an externally presented inducing stimulus is not necessary to trigger a photism and that simply activating the concept of a digit is sufficient.

As a measure of consistency, we asked C. to name the colour of her photisms elicited by the digits 1 to 9 shown in random order, 10 times each. C.'s pairings between digits and colour names were 100% consistent across repetitions.

To assess automaticity, we used a Stroop task. Colour-naming reaction times were recorded for coloured squares (baseline condition) or for digits displayed in colours that were either congruent or incongruent with C.'s photisms (Fig. 1). Her mean reaction times for each condition were: congruent, 552 ms (error, 1.4%); baseline, 545 ms (error, 0.0%); incongruent, 797 ms (error, 2.8%). An analysis of variance ($F(2,410) = 74.88$, $P < 0.001$) and Tukey post-hoc analyses revealed that incongruent trial reaction times were significantly slower than baseline ($P < 0.001$) and congruent trial reaction times ($P < 0.001$) — the last two conditions did not differ.

C.'s large significant difference in congruent/incongruent reaction times (245 ms) contrasts with the small, non-significant, congruent/incongruent differences (ranging from –11 to 17 ms) of eight non-synaesthetes tested using identical stimuli. Figure 1a shows the Stroop performance of C. and of the non-synaesthete most similar to C. in terms of baseline-trial colour-naming reaction times. C.'s performance indicates that the digits automatically elicited photisms that interfered with her colour-naming on incongruent trials.

For C., viewing the digit 7 automatically induces a yellow photism. To assess whether C.'s photisms could be elicited without physically showing her an inducing stimulus, we presented, in sequence, a digit (for example, 4), an operator (for example, +), a second digit (for example, 3), and a colour patch. C. had to name the colour patch as quickly as possible, and then report the solution to the arithmetic problem. On each trial, the colour patch was either

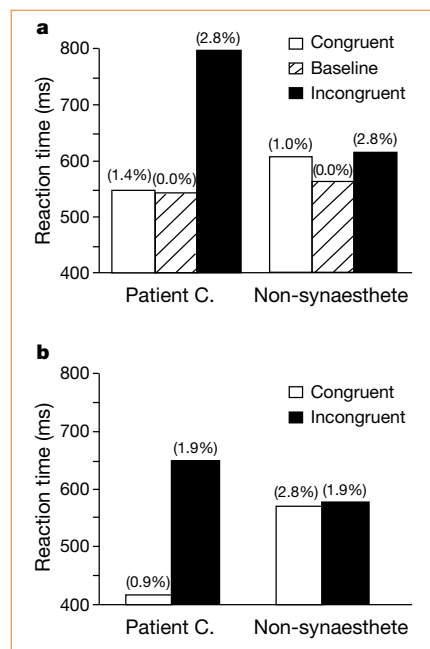


Figure 1 Colour-naming reaction times (and errors) for a digit–colour synaesthete, C., and for representative non-synaesthetes. **a**, Digit Stroop task: on each trial, reaction times were recorded for naming coloured squares ('baseline') or digits displayed in colours that were either congruent or incongruent with C.'s photism colours. There were 144 trials of each type, randomly intermixed. **b**, Arithmetic Stroop task: on each trial, participants were presented, in sequence, with a fixation cross, a digit, an arithmetic operator, a second digit, and a colour patch. Participants named the colour patch as quickly as possible and then reported the arithmetic solution. Colour patches were either congruent (108 trials) or incongruent (324 trials) with C.'s photism colours for the arithmetic solutions. Trial types were intermixed. Further details are available from the authors.

congruent or incongruent with the photism associated with the arithmetic solution. For example, $5 + 2$ was followed either by a yellow patch, which is congruent with C.'s photism colour for 7, or a red patch, which is incongruent with C.'s photism for 7. To discourage C. from using the arithmetic solution to predict the colour of the patch, we presented three times as many incongruent trials as congruent trials.

If automatic photisms can be induced simply by calculating the solution to arithmetic problems, then they should interfere with C.'s ability to name the colour of the patch on incongruent trials but not on congruent trials. Consistent with our reasoning, C.'s colour-naming reaction times were significantly slower on incongruent trials (mean, 650 ms) than on congruent trials (mean, 414 ms), $t(390) = 19.48$, $P < 0.001$. This large congruent/incongruent difference (236 ms) was 19.3 standard deviations away from the mean difference (1 ms) obtained for eight non-synaesthetes. Figure 1b shows C.'s performance compared with the performance of the non-synaesthete who was most similar to C. in terms of overall reaction times.

C.'s large difference in congruent/incongruent reaction times (236 ms) indicates that automatic photisms were induced by the arithmetic solution (for example, the yellow photism associated with the digit 7 was generated in response to calculating $5 + 2$). This suggests that an external stimulus (for example, a physically present numeral 7) is not required to trigger a

photism. Rather, activating the concept of a digit by a mental calculation was sufficient to induce a colour experience. Thus, although C.'s photisms are both consistent and automatic, they do not require a physical inducing stimulus to elicit them.

Stroop-type paradigms have shown that photisms are automatic and involuntary⁴. Brain imaging has demonstrated that photisms activate visual-association areas involving binding colour to form⁵. Cognitive research, brain imaging and even genetics (synaesthesia runs in families)⁶ all attest to the reality of synaesthesia. Although the neural architecture and neurodevelopmental aspects of synaesthesia are not yet fully understood, the finding that photisms can be concept-driven may help to guide our understanding of this rare phenomenon.

Mike J. Dixon, Daniel Smilek, Cera Cudahy, Philip M. Merikle

Department of Psychology, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
e-mail: mjdixon@watarts.uwaterloo.ca

- Baron-Cohen, S., Wyke, M. & Binnie, C. *Perception* **16**, 761–767 (1987).
- Mills, C. B., Boteler, E. H. & Oliver, G. K. *Cogn. Neuropsychol.* **16**, 181–191 (1999).
- Odgaard, E. C., Flowers, J. H. & Bradman, H. L. *Perception* **28**, 651–664 (1999).
- Cytowic, R. E. *The Man who Tasted Shapes* (Putnam, New York, 1993).
- Frith, C. D. & Paulsen, E. in *Synaesthesia: Classic and Contemporary Readings* (eds Baron-Cohen, S. & Harrison, J. E.) 123–147 (Blackwell, Cambridge, Massachusetts, 1997).
- Bailey, M. E. S. & Johnson, K. J. in *Synaesthesia: Classic and Contemporary Readings* (eds Baron-Cohen, S. & Harrison, J. E.) 182–210 (Blackwell, Cambridge, Massachusetts, 1997).

Avian phenology

Climate change and constraints on breeding

Although climate change apparently affects the breeding patterns of many animals^{1–3}, the wider implications for breeding success are unclear. Here we describe an energy trade-off between reproduction and maintenance that occurs during cold weather in great tits (*Parus major* L.), pointing to a thermal constraint on the timing of egg laying. Our observations indicate that the fine-scale pattern of climate change could be critical to the reproduction of some species and underlies previously unexplained variation in the breeding success of other temperate birds.

To predict the effects of climate change on populations and communities, we must understand the cues and constraints that determine the timing of breeding. Selection will favour birds whose chicks hatch when most food is available, and the earliest breeders are often the fittest^{4,5}. By measuring the energy expenditure of free-living birds during egg laying, however, we have found that this selection is opposed by temperature-dependent constraints on early laying. In combination with variable patterns of spring warming, these constraints can explain some puzzling results attributed to global warming^{2,6}.

We used the doubly labelled water technique⁷ to study the daily energy expenditure (DEE) of laying great tits in central Scotland during 1997 and 1998. DEE increased as ambient temperature fell (Fig. 1a), probably as a result of increased thermostatic requirements or temperature-dependent food availability. After controlling for temperature, high DEE was also associated with the production of large eggs in 1997 (Fig. 1b).

The association of high energy-expenditure with both low temperatures and the production of large eggs indicates that there are energetic constraints on the fitness benefits of early laying. Large eggs increase fitness because they produce heavier fledglings⁸, which in turn have a higher recruitment rate⁹, but low temperatures may hamper the formation of such eggs. In 1997, the largest eggs were associated with DEE (even when corrected to average temperatures) at or above four times the basal metabolic rate (BMR) ceiling for cost-free sustained metabolism^{9,10} (BMR, 29.59 kJ d^{−1}; ref. 11) (Fig. 1b).

DEE also approached this fitness-cost threshold when laying occurred at low temperatures (Fig. 1a). Consequently, in the absence of economies, laying large eggs at low temperatures would push DEE beyond the level at which fitness costs become evident^{9,10}. In these circumstances, we predict

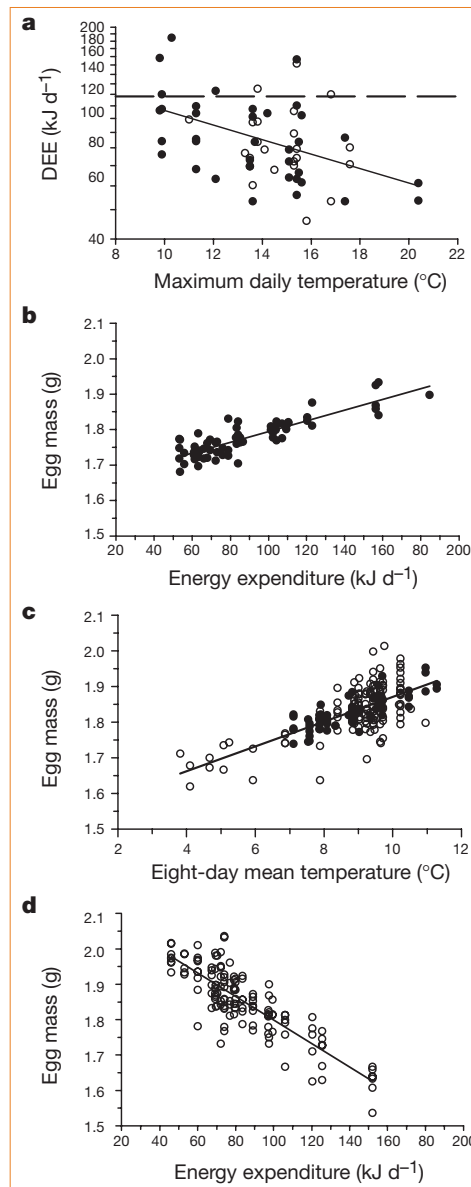


Figure 1 Variation in daily energy expenditure with ambient temperature and egg size. Filled circles are 1997 data points and circles are 1998 data points. **a**, Daily energy expenditure (DEE) and ambient temperature in 1997 ($n=36$) and 1998 ($n=23$) ($F_{1,57}=13.4$, $r^2=0.18$, $P=0.0006$). The dashed line indicates $4 \times$ basal metabolic rate. Energy expenditure was measured over a 24-hour period using doubly labelled water. **b**, Egg mass and DEE in 1997 (DEE, $\chi^2=0.3$, n.s.; year, $\chi^2=2.3$, n.s.; DEE $\times$ year interaction term, $\chi^2=14.8$, $P=0.0001$). The data are partial residuals, adjusted to the mean of the explanatory terms included in the analysis but not shown in the figures. Results were analysed in a linear mixed model using restricted maximum likelihood (REML)¹⁴ to account for the non-independence of eggs from the same clutch and the unbalanced structure of the data. Explanatory variables were laying date, energy expenditure, ambient temperature, year of study, clutch size and location within study area (body mass, tarsus size, age and egg position in laying sequence were non-significant: all $P>0.40$). **c**, Egg mass and temperature averaged over the eight days before the laying of each egg. The relation comes from the REML analysis in **b**, again plotted as partial residuals adjusted to the mean of the other explanatory variables in the model. Progressive lengthening of the temperature-averaging period from one day up to nine days produced a consistent decline in residual variance up to eight days, after which the residual increased (temperature, $\chi^2=18.4$, $P=0.00002$). **d**, Egg mass and DEE in 1998; relation comes from the REML analysis in **b**. 1998 was a poor breeding year, as indicated by later clutch initiation, smaller clutch size, and lower fledgling success (mean clutch size 1997, 7.5; 1998, 6.7; $t_{107}=2.38$, $P=0.009$; mean clutch initiation date in 1997 was 22 April, and 29 April in 1998, $t_{109}=270$, $P<0.0001$; mean fledgling success in 1997 = 89% and 68% in 1998; $F_{1,16}=5.91$, $P=0.027$).

that energy will be diverted from egg production, an idea that is supported by the observed decline in egg size with decreasing temperature during egg formation (Fig. 1c). Hence in some years, the risk of producing poor-quality eggs early in the season (when temperatures are unpredictable and normally lower) will tend to oppose the selection for earlier laying that has been repeatedly noted^{4–6}.

Our results provide a further explanation for reported changes in the breeding phenology of birds. If, as we argue, temperature is a constraint on laying date and not simply a cue for the timing of nestling food, we would predict that laying date should be more dependent on temperature in smaller species, as their thermoregulatory costs will be relatively higher and more sensitive to temperature¹². Using the available data on body mass¹³ and on laying advancement in response to increased temperature¹, we find that,

indeed, laying is advanced most in the smallest species (Spearman rank correlation, $r_s = -0.34$, $n=36$, $P=0.04$).

Our study also implies that the long-term responses of avian breeding to climate change may depend on the detailed patterning of spring temperatures. In a Dutch population of great tits, the laying date during 1973–95 did not advance, but the peak of nestling-food availability did⁶, a mismatch that led to reduced breeding success. By contrast, a population near Oxford, UK, showed a consistent advancement in laying and hatching dates, with no decline in breeding success².

The key difference between these studies is the pattern of temperature change. In the Netherlands, temperatures increased in late spring, but not at egg-laying time, advancing the development of caterpillars but not great-tit laying⁶. For the UK population, temperature consistently increased throughout spring, maintaining synchrony

between nestlings and their food (UK Meteorological Office, Oxford station; early period, 1 March–15 April; late period, 16 April–15 May; 0.08 °C increase per year, $F_{1,43} = 9.92$, $P = 0.003$; period $\times$ year interaction, $F_{1,42} = 0.02$ n.s.). If temperature acts as a cue to anticipate caterpillar emergence, selection may refine the use of this and other cues to resynchronize chicks with their food. This cannot happen, however, if temperature *per se* constrains laying date and egg size.

Our results indicate that important fitness trade-offs can operate during egg laying, but many questions remain. Under poor breeding conditions in 1998, for example, we found that the relation between energy expenditure and egg size was reversed (Fig. 1d). This negative relation cannot be causal, and we suggest that it is confounded by territory or phenotypic quality, or by changing resource allocation patterns under poor conditions. Although we demonstrate this remarkable switch in response between years, we cannot yet explain it, which calls for detailed investigation of this phase of breeding.

Our results highlight the potential importance of the egg-laying phase in determining the timing of avian breeding, and the relations we have identified must be explained if we are to understand responses to climate change. The use of coarse climate-change indices, such as spring temperature sum² or North Atlantic Oscillation³, can be inappropriate in such studies. Whereas analysis of temperature changes at this scale suggests that many populations can advance their timing without adverse effect², our results caution that complex changes in temperature patterns within seasons can disrupt breeding phenology.

Ian R. Stevenson, David M. Bryant

Institute of Biological Sciences, University of Stirling, Stirling FK9 4LA, UK
e-mail: i.r.stevenson@stir.ac.uk

1. Crick, H. Q. P. & Sparks, T. H. *Nature* **399**, 423–424 (1999).
2. McCleery, R. H. & Perrins, C. M. *Nature* **391**, 30–31 (1998).
3. Forchhammer, M. C., Post, E. & Stenseth, N. C. *Nature* **391**, 29–30 (1998).
4. Perrins, C. M. *Ibis* **138**, 2–15 (1996).
5. van Noordwijk, A. J., McCleery, R. H. & Perrins, C. M. *J. Anim. Ecol.* **64**, 451–458 (1995).
6. Visser, M. E., van Noordwijk, A. J., Tinbergen, J. M. & Lessells, C. M. *Proc. R. Soc. Lond. B* **265**, 1867–1870 (1998).
7. Speakman, J. R. *Doubly Labelled Water: Theory and Practice* (Chapman & Hall, London, 1997).
8. Both, C., Visser, M. E. & Verboven, N. *Proc. R. Soc. Lond. B* **266**, 465–469 (1999).
9. Drent, R. H. & Daan, S. *Ardea* **68**, 225–252 (1980).
10. Tinbergen, J. & Verhulst, S. J. *J. Anim. Ecol.* **69**, 323–334 (2000).
11. Hashim, R. B. *Basal and Free-living Energy Expenditure by Birds: Studies of Great Tits and Japanese Quail*. Thesis, Univ. Stirling (1996).
12. Reiss, M. J. *The Allometry of Growth and Reproduction* (Cambridge Univ. Press, Cambridge, 1989).
13. Cramp, S. et al. (eds) *Handbook of the Birds of Europe, the Middle East and North Africa: The Birds of the Western Palearctic* (Oxford Univ. Press, Oxford, 1977–95).
14. Patterson, H. D. & Thompson, R. *Biometrika* **58**, 545–554 (1971).

Signalling pathways

Kinase regulation in inflammatory response

The transcription factor NF- κ B is a pivotal regulator of innate immune responses, whose activity is rapidly induced by proinflammatory stimuli, most notably the tumour-necrosis factor TNF α and interleukin-1, viruses, and components of bacterial cell walls¹. In addition, NF- κ B protects cells from the induction of programmed cell death by pro-apoptotic stimuli such as TNF α (refs 2,3). Another important anti-apoptotic signal-transducing protein is the protein kinase Akt (also known as protein kinase B), whose activity is strongly stimulated by growth factors⁴. Ozes *et al.*⁵ have suggested that Akt is involved in the TNF α -mediated activation of NF- κ B, implying that some of the anti-apoptotic activity of Akt may be mediated through NF- κ B. However, we have failed to detect any involvement of Akt in the signalling pathway through which TNF α leads to NF- κ B activation.

NF- κ B activation requires the degradation of I κ B proteins, which otherwise keep NF- κ B inactive in the cytoplasm of non-stimulated cells¹. I κ B degradation and NF- κ B activation rely on the I κ B kinase (IKK), which consists of two catalytic subunits, IKK α and IKK β , and a regulatory subunit, IKK γ /NEMO¹. *In vitro*, both IKK α and IKK β phosphorylate I κ Bs at the sites that mediate their phosphorylation-dependent degradation¹. However, *in vivo* only IKK β , and not IKK α , is required for I κ B degradation in response to proinflammatory stimuli, including TNF α (refs 6–8). IKK β is also essential for preventing TNF α -induced apoptosis⁶. Furthermore, phosphorylation

of IKK β is required for activation of the entire IKK complex by proinflammatory stimuli, whereas phosphorylation of IKK α is not essential⁹.

Although the mechanism of IKK activation by proinflammatory stimuli is relatively well understood, the identity of physiological IKK kinases that mediate its activation in response to different stimuli is the subject of intense investigation¹⁰. We were therefore interested in the report by Ozes *et al.*⁵ that TNF α treatment of HEK293 or HeLa cells results in activation of Akt which subsequently phosphorylates IKK α at a threonine residue at position 23 (T23), a site claimed to be essential for IKK and NF- κ B activation⁵. Furthermore, an inhibitor of phosphatidylinositol-3-OH kinase (PI(3)K) that prevents Akt activation⁴ was shown to diminish NF- κ B activation by TNF α (ref. 5). As Ozes *et al.* did not directly examine changes in IKK activity⁵, we have investigated whether Akt could be a regulator of IKK catalytic activity.

Stimulation of HeLa cells with insulin-like growth factor (IGF)-1 results in phosphorylation of Akt at serine 473 (S473), an indicator of its activation⁴. No increase in Akt S473 phosphorylation was detected in cells incubated with TNF α for up to 30 min (Fig. 1a), a period sufficient for full IKK activation⁹. Whereas TNF α stimulated NF- κ B DNA-binding activity, I κ B α phosphorylation and degradation, and IKK activity, IGF-1 did not elicit any of these responses (Fig. 1a).

Activation of Akt by IGF-1 can be blocked by PI(3)K inhibitors such as LY294002 or wortmannin⁴. Neither wortmannin nor LY294002 inhibited activation of IKK by TNF α (Fig. 1b). Thus if PI(3)K and Akt are involved in stimulation of NF- κ B activity, they most probably act after

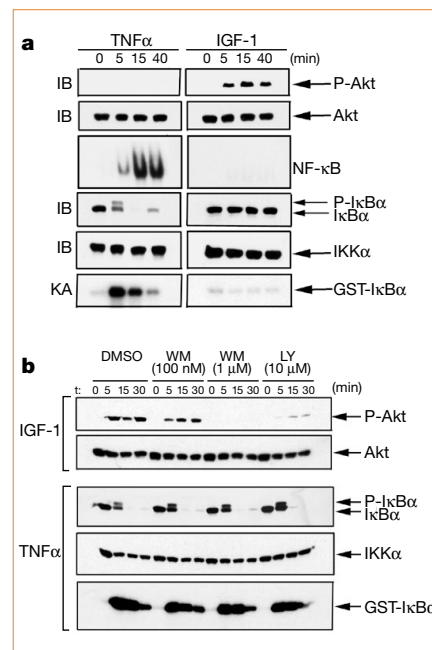


Figure 1 Akt is not involved in IKK activation. **a**, HeLa cells were serum-starved for 24 h then either left untreated or stimulated with TNF α (20 ng ml⁻¹) or IGF-1 (50 ng ml⁻¹) for the indicated times, then lysed. Top two panels, Akt activation was analysed by immunoblotting (IB) with antibody specific to phospho-Akt (S473). Total Akt was also determined. Third panel, NF- κ B DNA-binding activity was measured by electrophoretic mobility-shift assay. Fourth and fifth panels, I κ B α degradation was analysed by immunoblotting. I κ B α phosphorylation is indicated by the appearance of a slower-migrating species, denoted P-I κ B α . IKK α expression was determined by immunoblotting. Bottom panel, IKK activation measured by an immunocomplex (with GST, glutathione-S-transferase) kinase assay (KA). **b**, Inhibition of Akt activation does not interfere with activation of IKK. Serum-starved HeLa cells were pretreated for 30 min with dimethylsulphoxide (DMSO), wortmannin (WM; 100 nM or 1 μ M) or LY294002 (10 μ M). Cells were left untreated or stimulated with TNF α or IGF-1. At the indicated time points, activation of Akt, phosphorylation and degradation of I κ B α and activation of IKK were measured as in **a**.

NF- κ B has entered the nucleus, as has been proposed on the basis of genetic and biochemical analysis^{11,12}.

We also examined the claim that Akt activates IKK through phosphorylation of IKK α at T23. The results of Ozes *et al.* were obtained *in vitro*⁵, but we examined whether endogenous IKK α or IKK β are phosphorylated on threonine residues in TNF α -treated cells. In both non-stimulated or TNF α -stimulated cells, IKK α and IKK β were phosphorylated only at serine residues and no threonine phosphorylation could be detected (data not shown). In addition, overexpression of constitutively active Akt in HEK293 cells did not result in stimulation of IKK α -associated I κ B kinase activity (data not shown).

It should also be noted that IKK can be fully activated by TNF α or interleukin-1 in IKK α -deficient cells^{7,8}. Thus, IKK α phosphorylation by Akt or any other protein kinase is not essential for IKK activation. Although PI(3)K activation can potentiate NF- κ B activity in the nucleus by IKK-independent mechanisms^{11,12}, we find no evidence for the involvement of the PI(3)K–Akt signalling pathway in IKK activation, which is essential for sending NF- κ B to the nucleus¹.

Mireille Delhase, Nanxin Li, Michael Karin
Laboratory of Gene Regulation and Signal Transduction, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0636, USA
e-mail: karinoffice@ucsd.edu

1. Rothwarf, D. M. & Karin, M. The NF- κ B Activation Pathway: A Paradigm in Information Transfer from Membrane to Nucleus *www.stke.org/cgi/content/full/OC_sigtrans/1999/5/re1* (1999).
2. Beg, A. A. & Baltimore, D. *Science* **274**, 782–784 (1996).
3. Liu, Z.-G., Hu, H., Goeddel, D. V. & Karin, M. *Cell* **87**, 565–576 (1996).
4. Hemmings, B. A. *Science* **275**, 628–630 (1997).
5. Ozes, O. N. *et al.* *Nature* **401**, 82–85 (1999).
6. Li, Q., Van Antwerp, D., Mercurio, F., Lee, K.-F. & Verma, I. M. *Science* **284**, 321–325 (1999).
7. Hu, Y. *et al.* *Science* **284**, 316–320 (1999).
8. Takeda, K. *et al.* *Science* **284**, 313–316 (1999).
9. Delhase, M., Hayakawa, M., Chen, Y. & Karin, M. *Science* **284**, 309–313 (1999).
10. Tojima, Y. *et al.* *Nature* **404**, 778–782 (2000).
11. Sizemore, N., Leung, S. & Stark, G. R. *Mol. Cell. Biol.* **19**, 4798–4805 (1999).
12. Madrid, L. V. *et al.* *Mol. Cell. Biol.* **20**, 1626–1638 (2000).

Ozes *et al.* reply — Delhase *et al.* take issue with our claim¹ that Akt induces activation of NF- κ B by phosphorylating IKK α , contending that IKK α plays no role in the activation by TNF of NF- κ B, and consequently that Akt could not affect NF- κ B through IKK α . They point out that Hu *et al.*² have shown that cells deficient in IKK α have normal TNF-induced NF- κ B activity, but this has been refuted by Li *et al.*³, who reported significant reduction of TNF-induced NF- κ B in IKK α -deficient cells. Indeed, the observations of Hu *et al.*² show that degradation of I κ B α is diminished in cells from IKK α -deficient mice and are

therefore not consistent with the conclusion that IKK α plays no role in TNF induction of NF- κ B. Furthermore, deficiency of IKK β only partially impairs TNF-induced NF- κ B activation^{4,5}, which reserves a role for IKK α in this pathway. Others^{5–7} have shown that activation of the IKK complex is dependent on the kinase activity of IKK α to activate IKK β . Thus, strong evidence supports a role for IKK α in TNF induction of NF- κ B.

Delhase *et al.* only tested the role of Akt on NF- κ B activation in HeLa cells in which they did not observe activation of Akt by TNF. As the involvement of inflammatory stimuli, including TNF, TRAF-6, IL-1 and LPS^{1,8–12} in PI(3)K/Akt activation is well documented, Delhase *et al.* should have investigated the Akt/NF- κ B connection in some of these systems.

The properties of HeLa cells are known to vary greatly among different laboratories, and the inability of Delhase *et al.* to detect Akt activation in a particular line of HeLa cells is not only inconsistent with our observations¹ but also with others^{13,14}. In view of our results, it is interesting that NF- κ B reporter gene activity in HeLa cells can be induced by constitutively active Akt and that this is inhibited by dominant-negative IKK β (ref. 15). We have done experiments in seven cell types and found that the amount of Akt, the extent to which it is activated by TNF, and its effects on IKK α are cell-type specific.

PI(3)K/Akt signalling has been implicated in NF- κ B activation induced by inflammatory^{1,8,11}, mitogenic¹⁶ and oncogenic stimuli¹⁷, as well as by T-cell activation¹⁷ and signalling through G-protein receptors¹⁵. Significant mechanistic differences exist as to how Akt is incorporated into NF- κ B signalling, with the most unsettled issue being the role of the I κ B kinase in NF- κ B activation. Several groups^{1,9,16} have reported direct physical and functional interactions between Akt and IKK, and as the only established function of IKK is to phosphorylate I κ B, this suggests that I κ B is a target for Akt^{1,9,16,18}.

Others have found no evidence for the involvement of Akt in I κ B degradation, but rather propose an I κ B-independent mechanism in which Akt affects the transcriptional activity of NF- κ B (refs 8,11,17). However, even those in favour of an I κ B-independent mechanism acknowledge the requirement for IKK activation for Akt effects on NF- κ B transactivation¹⁷. This raises the question of why, if IKK activity is required for regulation of the transactivation function of NF- κ B, activated IKK has stopped acting on I κ B, its preferred substrate.

These disparate observations point to deficiencies in our understanding of NF- κ B activation, and also suggest that the Akt/NF- κ B connection is cell-type- and

stimulus-specific. Consistent with this idea, constitutively active Akt alone induces NF- κ B reporter gene activity in some cells^{1,11,15,17}, whereas in others, signals from other pathways are required for Akt to manifest its effect^{8,18}. Likewise, activation of NF- κ B by TNF is mediated by a PI(3)K/Akt pathway in some cell types^{1,8} but not others^{12,16}. We believe that the attempt of Delhase *et al.* to define a complex and incompletely understood signalling system based on experiments with a single cell type is an example of misplaced reductionism.

Osman Nidai Ozes*, Lindsey D. Mayo*, Jason A. Gustin*, Susan R. Pfeffer†, Lawrence M. Pfeffer†, David B. Donner*

*Department of Microbiology and Immunology, Indiana University School of Medicine and the Walther Oncology Center, Indianapolis, Indiana 46202, USA

†Department of Pathology, University of Tennessee Health Science Center, Memphis, Tennessee 38163, USA

1. Ozes, O. N. *et al.* *Nature* **401**, 82–85 (1999).
2. Hu, Y. *et al.* *Science* **284**, 316–320 (1999).
3. Li, Q. *et al.* *Genes Dev.* **13**, 1322–1328 (1999).
4. Li, Q. *et al.* *Science* **284**, 321–325 (1999).
5. Tanaka, M. *et al.* *Immunity* **10**, 421–429 (1999).
6. O'Mahony, A. *et al.* *Mol. Cell. Biol.* **20**, 1170–1178 (2000).
7. Yamamoto, Y., Yin, M.-J. & Gaynor, R. B. *Mol. Cell. Biol.* **20**, 3655–3666 (2000).
8. Reddy, S. A. G. *et al.* *J. Immunol.* **164**, 1355–1363 (2000).
9. Zhou, B. P. *et al.* *J. Biol. Chem.* **275**, 8027–8031 (2000).
10. Wong, B. R. *et al.* *Mol. Cell. Biol.* **19**, 1041–1049 (1999).
11. Sizemore, N. *et al.* *Mol. Cell. Biol.* **19**, 4798–4805 (1999).
12. Madge, L. A. & Pober, J. S. *J. Biol. Chem.* (in the press).
13. Pastorino, J. G. *et al.* *J. Biol. Chem.* **274**, 19411–19416 (1999).
14. Berra, E. *et al.* *J. Biol. Chem.* **273**, 10792–10797 (1998).
15. Xie, P. *et al.* *J. Biol. Chem.* (in the press).
16. Romashkova, J. A. & Makarov, S. S. *Nature* **401**, 86–90 (1999).
17. Madrid, L. V. *et al.* *Mol. Cell. Biol.* **20**, 1626–1638 (2000).
18. Kane, L. P. *et al.* *Curr. Biol.* **9**, 601–604 (1999).

Correction

Energy for microbial life on Europa

C. Chyba

Nature **403**, 381–382 (2000)

In this communication, I estimated the biomass that could be supported by mixing the ice irradiation products HCHO and H₂O₂ into Europa's ocean, finding values that ranged from ~10⁷ g to ~10¹⁰ g. However, a calculational error in the higher estimate has recently come to my attention (E. J. Gaidos, K. N. Nealson and J. L. Kirschvink, personal communication). The correct result is ~10⁹ g, making detectability of potential European life under this model considerably more difficult. I present a revised calculation and discuss related practical issues in the search for life on Europa at http://www.seti.org/pdf/chyba_lander.pdf.

Erratum

Do cockroaches 'know' about fluid dynamics?

D. Rinberg, H. Davidowitz

Nature **405**, 756 (2000)

The units of the r.m.s. of the square of the wind velocity shown in the x-axis label of Fig. 1a should be cm² s⁻², and not m² s⁻² as published.

Fringe is a glycosyltransferase that modifies Notch

Daniel J. Moloney^{*†}, Vladislav M. Panin^{†‡}, Stuart H. Johnston^{†§||}, Jihua Chen[¶], Li Shao^{*}, Richa Wilson[‡], Yang Wang[#], Pamela Stanley[¶], Kenneth D. Irvine[‡], Robert S. Haltiwanger^{*} & Thomas F. Vogt^{§||}

^{*} Department of Biochemistry and Cell Biology, Institute for Cell and Developmental Biology, SUNY-Stony Brook, Stony Brook, New York 11794-5215, USA

[‡] Waksman Institute and Department of Molecular Biology and Biochemistry, Rutgers, The State University, Piscataway, New Jersey 08854, USA

[§] Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

[¶] Department of Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, New York, New York 10461, USA

[#] Department of Pharmacokinetics and Metabolism, Genentech, Inc., South San Francisco, California 94080, USA

[†] These authors contributed equally to this work

Notch receptors function in highly conserved intercellular signalling pathways that direct cell-fate decisions, proliferation and apoptosis in metazoans. Fringe proteins can positively and negatively modulate the ability of Notch ligands to activate the Notch receptor. Here we establish the biochemical mechanism of Fringe action. *Drosophila* and mammalian Fringe proteins possess a fucose-specific β 1,3 *N*-acetylglucosaminyltransferase activity that initiates elongation of *O*-linked fucose residues attached to epidermal growth factor-like sequence repeats of Notch. We obtained biological evidence that Fringe-dependent elongation of *O*-linked fucose on Notch modulates Notch signalling by using co-culture assays in mammalian cells and by expression of an enzymatically inactive Fringe mutant in *Drosophila*. The post-translational modification of Notch by Fringe represents a striking example of modulation of a signalling event by differential receptor glycosylation and identifies a mechanism that is likely to be relevant to other signalling pathways.

The Notch receptors have a central role in the highly conserved and widely used signal transduction pathways that influence numerous cell-fate decisions in metazoans¹. Deregulation of Notch receptor signalling results in a wide variety of developmental defects in both invertebrates and vertebrates, and is associated with abnormalities of tissue organization and cancer in humans². Two families of conserved ligands, Delta and Serrate/Jagged, bind to and activate Notch receptors. Both the Notch receptors and their ligands are large, single-pass transmembrane proteins that include in their extracellular domains several tandem epidermal growth factor (EGF)-like repeats. Proteins that modulate the strength of Notch signalling are being identified at all levels of the Notch signal transduction cascade³. One modulator of particular interest is Fringe, a secreted protein that both positively and negatively modulates the ability of Notch ligands to activate Notch signalling^{4–6}.

Drosophila Fringe (D-Fng) was first identified because of its role in positioning Notch activation in specialized boundary cells along the dorsal–ventral compartment border of the wing⁷. Subsequently, D-Fng and its three vertebrate homologues, Lunatic Fringe (*Lfng*), Manic Fringe (*Mfng*) and Radical Fringe (*Rfng*), were shown to have a similar role in positioning Notch signalling in many tissues of both *Drosophila* and vertebrates^{8–11}. Genetic studies in *Drosophila* indicate that Fringe acts cell-autonomously to inhibit Serrate/Jagged-dependent Notch activation and to potentiate Delta-dependent Notch activation^{4–6}. Fringe is thought to act upstream of receptor activation because a constitutively activated form of Notch appears to be insensitive to Fringe⁴; however, the molecular mechanism by which Fringe modulates Notch signalling has not been determined.

Fringe proteins share a weak sequence similarity with a class of bacterial glycosyltransferases¹², and recent studies of mammalian Notch1 revealed two unusual forms of glycosylation on the EGF repeats of its extracellular domain¹³. Both *O*-linked fucose and *O*-linked glucose are present on Notch1 as monosaccharides and in extended forms of up to tri- and tetrasaccharides¹³. These forms of glycosylation are unusual in the direct linkage of fucose or glucose to

protein and in the elongation of fucose^{14,15}. Here we show that *Drosophila* and mammalian Fringe proteins possess a fucose-specific β 1,3 *N*-acetylglucosaminyltransferase activity that catalyses the elongation of *O*-linked fucose on the EGF repeats of Notch, and that this enzymatic activity is essential for Fringe function *in vivo*.

Fringe elongates *O*-linked fucose on Notch

To investigate whether Fringe could alter either the *O*-linked fucose or the *O*-linked glucose structures on Notch1, we transfected mammalian Fringe into Chinese hamster ovary (CHO) cells and analysed the structures of these glycans. To facilitate the radiolabelling of the *O*-linked fucose structures, we used Lec1 cells, a CHO cell line that cannot synthesize complex or hybrid-type *N*-glycans^{16,17} — two large classes of glycans that incorporate fucose. CHO cells express Notch1 (ref. 13), but northern analyses did not detect any of the known mammalian Fringe gene transcripts. [³H]fucose-labelled *O*-linked saccharides were released from Notch1 that was isolated from Lec1 cells stably transfected with either a complementary DNA expression plasmid encoding a Mfng–alkaline phosphatase (Mfng–AP) fusion protein or a control plasmid (Fig. 1a). Both Mfng and Mfng–AP modulate Notch signalling when ectopically expressed in *Drosophila* (refs 10, 11; and data not shown).

Analysis of *O*-linked glucose structures from Notch1 revealed no difference in the spectrum of *O*-linked glucose saccharides produced in the presence or absence of Mfng–AP (data not shown). In contrast, expression of Mfng–AP caused a marked increase in the relative proportion of elongated *O*-linked fucose species on Notch1 (Fig. 1a). Three *O*-linked fucose-containing structures larger than the monosaccharide were observed: the disaccharide (GlcNAc- β 1,3-fucitol), trisaccharide (Gal- β 1,4-GlcNAc- β 1,3-fucitol), and tetrasaccharide (Sia- α 2,3-Gal- β 1,4-GlcNAc- β 1,3-fucitol) (Fig. 1a)¹³. We confirmed the β 1,3 linkage between *N*-acetylglucosamine (GlcNAc) and fucitol by showing that the disaccharide co-migrates with a synthetic GlcNAc- β 1,3-fucitol standard (Fig. 1c, d). Our observation that all three structures increased in size relative to the monosaccharide (Fig. 1a) suggested that Fringe might be catalysing the GlcNAc transfer that initiates elongation of *O*-linked fucose.

|| Present addresses: Exelixis, Inc., South San Francisco, CA 94080, USA (S.H.J.); Merck Research Laboratories, West Point, PA 19486-0004, USA (T.F.V.).

We observed an even more marked effect on *O*-linked fucose elongation on a fragment of the mouse Notch1 extracellular domain (EGF repeats 19–23) secreted from Lec1 cells expressing Mfng–AP (Fig. 1b). This fragment contains a single, highly conserved consensus site for *O*-linked fucose addition within EGF 20 (ref. 13). In the absence of Mfng–AP (Fig. 1b, top panel), *O*-linked fucose on EGF 19–23 was almost completely in the monosaccharide form; however, a marked increase in the tetrasaccharide form of *O*-linked fucose relative to the monosaccharide was observed in Mfng–AP expressing cells (Fig. 1b, bottom panel). The lack of a significant amount of di- or trisaccharide on the Notch1 EGF 19–23 fragment is probably because we analysed only the mature, secreted form. By contrast, endogenous Notch1 (Fig. 1a) was immunoprecipitated from cell lysates and therefore contained both biosynthetic intermediates and mature *O*-linked fucose structures. Alternatively, it is possible that different EGF repeats on Notch1 are elongated to different extents. The structure of the tetrasaccharide form of *O*-linked fucose on Notch1 EGF 19–23 was confirmed by exoglycosidase digestions and co-migration of the disaccharide with

standards (data not shown; and Fig. 1c, d). Thus, the presence of Mfng–AP leads to an increase in the elongation of *O*-linked fucose on the EGF repeats of Notch1 in CHO cells, and EGF 19–23 contains at least one *O*-linked fucose site that can be elongated.

Fringe is a GlcNAc-transferase

To determine whether Fringe is itself a fucose-specific β 1,3 *N*-acetylglucosaminyltransferase (β 1,3-GlcNAc-transferase), Fc- or His₆-tagged versions of Lfng, Mfng and D-Fng were expressed, affinity purified from medium and used in *in vitro* glycosyltransferase assays. The low molecular weight substrate *p*-nitrophenyl- α -L-fucose (pNp-fucose) structurally mimics *O*-linked fucose, and has been previously used to assay for an UDP-glucose: *O*-linked fucose β 1,3-glucosyltransferase¹⁸. Upon incubation of Lfng–Fc, Mfng–Fc or D-Fng–His with pNp-fucose and UDP-[³H]GlcNAc (the typical donor substrate for a GlcNAc-transferase), we observed significant GlcNAc-transferase activity (Fig. 2a, Table 1). A recombinant EGF repeat from the serum protein factor VII modified with *O*-linked fucose¹⁹ was also examined as a more physiological substrate. Factor VII EGF-*O*-fucose served as a specific acceptor in GlcNAc-transferase assays with all three Fringes (Fig. 2a, Table 1). Notably, product formation was completely dependent on the presence of the *O*-linked fucose on the EGF repeat, as EGF alone was not an acceptor (Fig. 2a, Table 1).

To specifically examine Notch as an acceptor, the extracellular domain of *Drosophila* Notch (ECN) bearing an amino-terminal Flag epitope (Flag–ECN) was expressed and affinity purified from S2 cells (Fig. 2b). We incubated Flag–ECN with UDP-[³H]GlcNAc and extracts of S2 cells expressing a Myc epitope-tagged D-Fng (Fng–Myc) or a control extract (Fig. 2b). Transfer of [³H]GlcNAc from UDP-[³H]GlcNAc to Flag–ECN occurred only in extracts with D-Fng–Myc. This activity was also exhibited by affinity purified D-Fng (data not shown). Fringe is therefore a GlcNAc-transferase that modifies the *O*-linked fucose glycans on EGF repeats, and

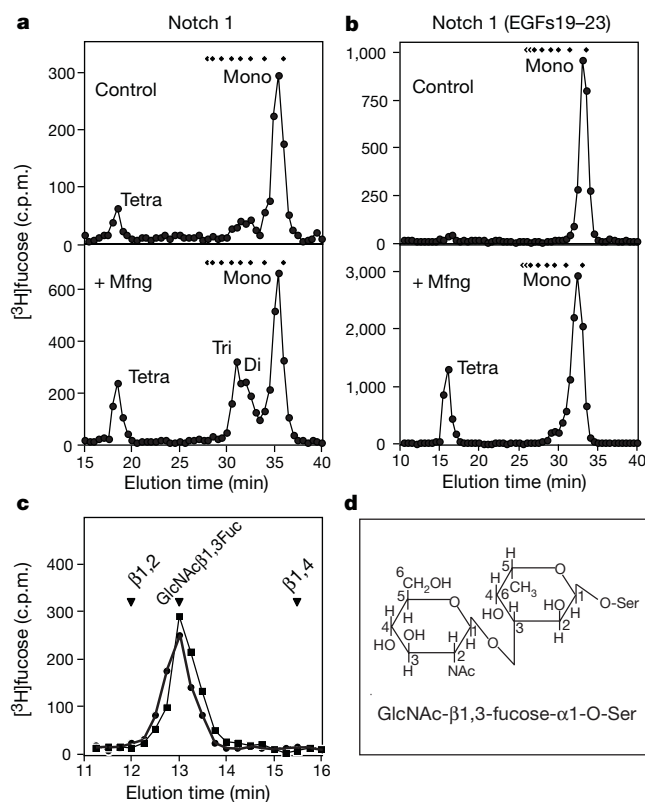


Figure 1 Manic Fringe elongates the *O*-linked fucose saccharides on full-length Notch1 and Notch1 EGF repeats 19–23. **a**, **b**, [³H]fucose-labelled *O*-linked saccharides were released by β -elimination from immunoprecipitated full-length, endogenous Notch1 (**a**) or from secreted Notch1 fragment EGF 19–23 (**b**) obtained from Lec1 CHO cells stably transfected with either a control plasmid (top panels) or with pMIRBMfng–AP (clone E11, bottom panels). Released oligosaccharides were analysed by gel-filtration chromatography with glucose size standards (diamonds), and structurally characterized as described¹³. Mono, fucitol; Di, GlcNAc- β 1,3-fucitol; Tri, Gal- β 1,4-GlcNAc- β 1,3-fucitol; Tetra, Sia- α 2,3-Gal- β 1,4-GlcNAc- β 1,3-fucitol. The overall increase in radioactivity seen with endogenous Notch in Mfng–AP cells (**a**) may reflect a better recovery of extracellular domain for technical reasons or be a consequence of Fringe-mediated protection from proteolysis (see Fig. 6c, iii). Mfng–AP cells express more EGF 19–23 (**b**, bottom panel) than the control cells (**b**, top panel). **c**, Product from Notch1 (circles) and Notch1 fragment EGF 19–23 (squares) isolated from cells expressing Mfng–AP were analysed by HPAEC chromatography as described¹³. Elution positions of synthetic standards in which GlcNAc is linked β 1,2, β 1,3 or β 1,4 to fucitol are shown (arrowheads). **d**, Structure of the disaccharide linked to serine (GlcNAc- β 1,3-fucose-*O*-Ser).

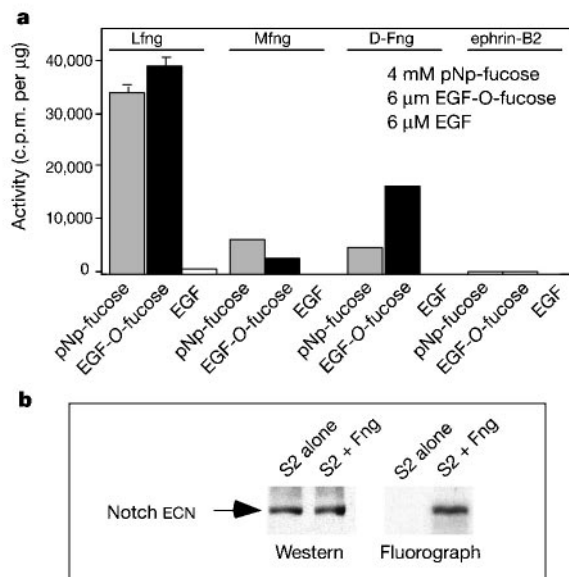


Figure 2 Luntatic, Manic and *Drosophila* Fringe catalyse the transfer of GlcNAc from UDP-[³H]GlcNAc to fucose *in vitro*. **a**, GlcNAc-transferase assays were performed using either pNp-fucose (4 mM) or factor VII EGF repeat (6 μ M) with (EGF-*O*-fucose) and without (EGF) *O*-linked fucose as acceptors. Assays were performed in duplicate with Fringe or control (ephrin-B2) proteins captured on beads. Error bars represent the range of duplicates. **b**, Affinity-purified Flag–ECN was incubated with UDP-[³H]GlcNAc and lysates of S2 cells stably expressing transfected D-Fng (S2 + Fng) or a control lysate (S2 alone). Equal portions of each sample were analysed by western blot with anti-Flag antibody (left panel), and by SDS–PAGE followed by fluorography (right panel).

Table 1 Lunatic, Manic and *Drosophila* Fringe are fucose-specific GlcNAc-transferases

Substrate	Enzyme				
	Lfng	Mfng	D-Fng (c.p.m. per μ g protein)	D-Fng (DEE)	ephrin-B2
UDP-GlcNAc as donor					
pNp-fucose (4 mM)	34,119	5,892	4,747	n.d.	< 500
EGF-O-fucose (6 μ M)	38,754	2,488	16,349	< 500	< 500
EGF (6 μ M)	< 500	< 500	< 500	n.d.	< 500
pNp-fucose (4 mM) + EDTA	< 500	< 500	< 500	n.d.	n.d.
pNp-galactose (4 mM)	< 500	< 500	< 500	n.d.	n.d.
pNp-glucose (4 mM)	< 500	< 500	< 500	n.d.	n.d.
UDP-galactose as donor					
pNp-fucose (4 mM)	< 500	< 500	< 500	n.d.	n.d.
UDP-glucose as donor					
pNp-fucose (4 mM)	< 500	< 500	< 500	n.d.	n.d.
UDP-GalNAc as donor					
EGF-O-fucose (6 μ M)	n.d.	n.d.	< 500	n.d.	n.d.

GlcNAc-transferase assays were performed in duplicate with the Fringe or control (ephrin-B2) proteins captured on beads (see Methods). The activity was calculated as the difference between product formed with and without acceptor substrate and normalized to the amount of Fringe or control protein. If the amount of product formed was less than the error (typically less than 500 c.p.m.) in the duplicate samples, the result was determined to be statistically insignificant and is reported as '< 500'. n.d., not determined.

sites for Fringe modification exist on the extracellular domain of Notch.

For all three Fringe proteins, the GlcNAc-transferase activity was specific for pNp-fucose as acceptor (pNp-galactose and pNp-glucose were not acceptors) and for UDP-GlcNAc as donor (UDP-galactose, UDP-glucose and UDP-GalNAc were not donors) (Table 1). As for most other glycosyltransferases²⁰, removal of manganese by EDTA inhibited Fringe activity (Table 1). Product formation was dependent on both Fringe protein and substrate concentrations (Fig. 3a–c; and data not shown). Under these assay conditions, Lfng–Fc had a higher specific activity than either D-Fng–His or Mfng–Fc (Table 1). Notably, EGF-O-fucose was a much better substrate than pNp-fucose, as equivalent transfer was obtained with 1000-fold less EGF-O-fucose (compare Fig. 3c and d, and Table 1). This result argues that the EGF repeat makes important contributions to substrate recognition by Fringe.

Product analysis showed that all three Fringes catalyse the formation of a β 1,3 linkage between GlcNAc and fucose (Figs 3e and 1d). Therefore, D-Fng, Lfng and Mfng are all fucose-specific β 1,3 *N*-acetylglucosaminyltransferases. This highly unusual glycosidic linkage (GlcNAc- β 1,3-fucose) has only been observed previously on EGF repeats^{13,14}, which indicates that Fringe may be directly responsible for causing elongation of *O*-linked fucose residues on the Notch extracellular domain (Fig. 1, 2b).

Fringe modulates Notch signalling in CHO cells

We adapted an established ligand-dependent Notch activation cell-culture assay to CHO cells^{21,22} to determine whether the elongation of *O*-linked fucose glycans on Notch1 correlates with the biological activity of Fringe. To monitor Notch activation, CHO cells were transiently transfected with a CBF1-luciferase reporter plasmid²³, and Notch was activated by co-culture of CHO cells with L cells expressing the Notch ligand Jagged1 (refs 21, 22). Jagged1-L cells caused a roughly fourfold induction of Notch signalling in wild-type CHO cells compared with control L cells (Fig. 4a). By contrast, Jagged1-dependent Notch activation was significantly reduced in CHO cells stably transfected with either Mfng–AP or Lfng–AP (Fig. 4a). Therefore, this CHO/L cell co-culture assay faithfully replicated both Jagged1-dependent Notch activation²¹ and the inhibitory effect of Fringe^{4–6,22}.

We obtained evidence that Fringe was acting directly on the *O*-linked fucose glycans of Notch1 to alter Notch signalling by carrying out the same experiments with CHO glycosylation mutants. Lec1 CHO cells completely lack complex and hybrid-type *N*-glycans on glycoproteins^{16,17}. Both Jagged1-dependent activation of Notch and inhibition of Notch activation by Mfng–AP were essentially identical in Lec1 and wild-type CHO cells (Fig. 4a).

Therefore, neither complex nor hybrid-type *N*-glycans are required for Jagged1-dependent activation of Notch, nor do these *N*-glycans have a significant role in mediating the effects of Fringe.

The fact that Mfng modulated Notch signalling in Lec1 cells is consistent with the observation that Mfng elongates *O*-linked fucose residues (Figs 1–3). However, Mfng should not be able to

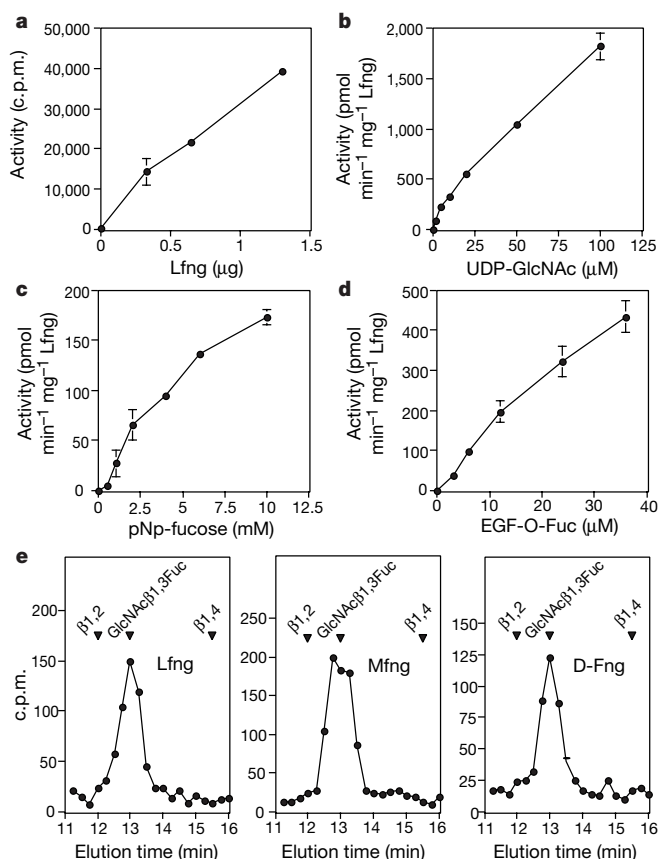


Figure 3 Lunatic, Manic and *Drosophila* Fringe are fucose-specific β 1,3 GlcNAc-transferases. GlcNAc-transferase assays were linear with respect to enzyme (Lfng–Fc, **a**) and substrates (UDP-GlcNAc, **b**; pNp-fucose, **c**; and EGF-O-fucose, **d**). Mfng and D-Fng proteins gave similar results (not shown). Assays were performed in duplicate using 2 μ M UDP-[³H]GlcNAc (**a**, **c**, **d**), 0.33 μ g Lfng–Fc (**b**–**d**) and 4 mM pNp-fucose (**a**, **b**). Error bars represent the range of duplicates. **e**, Product analysis was carried out as in Fig. 1. Identical results were obtained with pNp-fucose (shown) or EGF-O-fucose (not shown) as acceptors.

act on Notch in a cell that cannot add *O*-linked fucose to Notch. Lec13 CHO cells are defective in GDP-D-mannose-4,6-dehydratase^{24–27} (a key enzyme in the biosynthesis of GDP-fucose) and exhibit minimal transfer of fucose to any glycoconjugates²⁵. Notably, Jagged1-dependent activation of Notch was reduced in Lec13 cells (Fig. 4b). Expression of Mfng-AP in these cells did not cause any further decrease in Notch activation (Fig. 4b). The presence of a salvage pathway for GDP-fucose biosynthesis allows the Lec13 defect to be at least partially rescued by addition of fucose to the medium²⁴. Jagged1-dependent Notch activation and Fringe-dependent inhibition were partially restored in Lec13 cells supplemented with fucose (Fig. 4b). The combined data indicate that *O*-linked fucose modifications may be required for both Jagged1-dependent Notch activation and for the inhibition of Jagged1 signalling by Fringe.

D-Fng enzymatic activity is required *in vivo*

We generated a site-specific mutation in D-Fng, D236-D237-D238 to D236-E237-E238 (D-Fng(DEE)), to show that the GlcNAc-transferase activity of Fringe is essential for it to modulate Notch signalling *in vivo*. This aspartic acid patch is highly conserved among galactosyltransferases and several GlcNAc-transferases^{28,29}. Conservative changes within the DDD motif disrupt the catalytic activity of these enzymes without destabilizing their overall structure^{28,29}. Indeed, affinity-purified D-Fng(DEE) was enzymatically inactive (Table 1). To confirm that D-Fng(DEE) was synthesized and secreted normally, we co-expressed epitope-tagged wild-type, D-Fng-Myc, and DEE mutant, D-Fng(DEE)-HA, Fringe in *Drosophila* under the control of endogenous *fng* gene regulatory sequences by combining *UAS-fng* and *fng-Gal4* transgenes. Neither the Myc nor the HA (haemagglutinin A) tags have a detectable influence on *fng* expression or function (Fig. 5e, g)⁴. Wild-type and DEE mutant Fringe proteins were expressed at similar levels and with similar subcellular distributions (Fig. 5a–c).

We examined the biological activity of *fngDEE* by using an ectopic expression assay^{30,31}. The influence of *fng* on Notch signalling results in establishment of a stripe of Notch activation along the borders of *fng* expression^{4,9,30}. Ectopic expression of wild-type Fringe proteins under the control of *patched* regulatory sequences (*UAS-fng ptc-Gal4*) eliminates the normal Fringe expression border in the middle of the wing and induces an ectopic Fringe expression border in ventral cells. This results in corresponding changes in the pattern of Notch activation, as judged by Wingless expression in developing

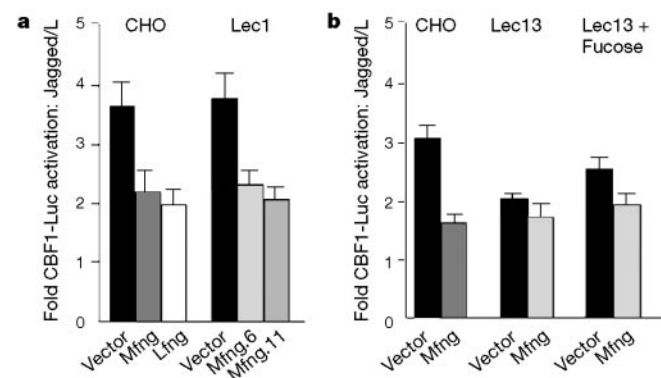


Figure 4 *O*-linked fucose is required for Jagged1-dependent activation of Notch and Fringe inhibition of CBF1 activation. **a**, Jagged1-dependent Notch activation is reduced in wild-type CHO cells stably expressing Mfng-AP ($n = 4$) or Lfng-AP ($n = 4$) and in Lec13 cells stably expressing Mfng-AP (clone F6, $n = 8$; clone E11, $n = 8$). **b**, Jagged1-dependent Notch activation is less in Lec13 CHO cells ($n = 4$) compared with in CHO cells ($n = 4$). In medium containing 1 mM fucose, Notch signalling by Jagged1 L cells increased in Lec13 cells ($n = 6$) and was inhibited in Lec13 cells expressing Mfng-AP ($n = 6$; $P < 0.03$). Error bars represent standard error of the mean.

wing imaginal discs (Fig. 5e) and by formation of margin bristles in adult wings (Fig. 5g)^{4,30}. Strikingly, ectopic expression of *fngDEE* failed to cause any discernible effect on Notch activation in the wing (Fig. 5d, f), or on any other aspects of *Drosophila* development (data not shown). Thus, the GlcNAc-transferase activity of Fringe is essential for it to modulate Notch signalling.

Discussion

Fringe proteins possess a new glycosyltransferase activity: they catalyse the transfer of GlcNAc in a β 1,3-linkage to *O*-linked fucose residues within EGF repeats (Fig. 6a). The transfer of GlcNAc to *O*-linked fucose may be the committed step in a pathway leading to the synthesis of a tetrasaccharide that terminates in a negatively charged sialic acid (Fig. 6a). Although NMR studies have shown that *O*-linked fucose has little effect on the tertiary structure of an EGF repeat¹⁹, elongation of *O*-linked fucose might well alter its intra- or intermolecular interactions. On the basis of our observations and of other studies that have revealed an autonomous mode of action for Fringe in cells responding to Notch ligands^{4,10,22}, we propose that Fringe modulates Notch signalling by direct

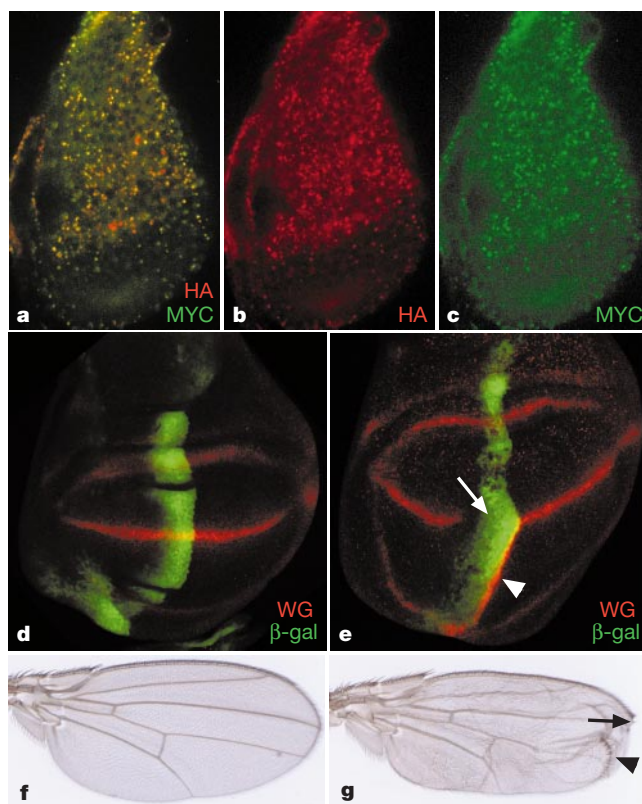


Figure 5 Mutation of a conserved acidic patch in Fringe renders it biologically inactive. **a–e**, *Drosophila* wing imaginal discs, with dorsal up; **f, g**, adult *Drosophila* wings. **a–c**, *UAS-fngDEE:HA UAS-fng:Myc fng-Gal4*, stained for expression of a wild-type tagged D-Fng protein (Myc, green) and D-Fng(DEE) tagged protein (HA, red). **b** and **c** show individual channels of the double stain (**a**). **d, e**, Discs stained for expression of Wingless (WG, red) and β -galactosidase (β -gal, green). WG expression is a marker of Notch activation⁹. β -gal marks the *patched* (*ptc*) expression stripe in which these FNG proteins are being ectopically expressed. **d**, *UAS-fngDEE:HA UAS-lacZ ptc-Gal4*; WG expression is indistinguishable from wild type. **e**, *UAS-fng:HA UAS-lacZ ptc-Gal4*; WG expression is inhibited where the *ptc* expression stripe intersects the dorsal–ventral boundary (arrow), and induced ectopically in ventral cells along the posterior edge of this stripe (arrowhead)^{4,30}. **f**, *UAS-fngDEE:HA ptc-Gal4*; the phenotype is indistinguishable from wild type. **g**, *UAS-fng:HA ptc-Gal4*; development of wing margin bristles is inhibited where the *ptc* expression stripe intersects the dorsal–ventral boundary (arrow), and induced ectopically in ventral cells along the posterior edge of this stripe (arrowhead)^{4,30}.

modification of *O*-linked fucose residues within EGF repeats of the Notch extracellular domain. Although Fringe and Notch must associate transiently in an enzyme–substrate complex, our results are inconsistent with the recent proposal that *Drosophila* Fringe functions by forming a stable complex with Notch³². We also note that we do not observe significant co-localization between Fringe and Notch *in vivo* (V.P. and K.I., unpublished data).

The ability to modulate *Drosophila* Notch signalling *in vivo* is conserved by mammalian *Fringe* genes^{10,11}. In addition, several of the consensus sites for the addition of *O*-linked fucose are conserved among *Drosophila* and vertebrate Notch proteins (EGF repeats 3, 20, 24, 26 and 31, Fig. 6b)¹³. Notably, none of these sites is near EGF repeats 11 and 12, which are necessary and sufficient for ligand binding in cell aggregation assays (Fig. 6b)³³. This lack of overlap suggests that Fringe does not affect Notch signalling by directly influencing binding between Notch and ligands presented on other cells. Consistent with this view, Fringe does not alter Notch–ligand binding in *Drosophila* S2 cell aggregation assays (ref. 34; V.P. and

K.I., unpublished data), and its influence on Notch1 activation does not correlate with a simple effect on ligand binding in mammalian co-culture assays²². As forms of Notch lacking an extracellular domain signal constitutively and are insensitive to Fringe^{4,22}, we propose that elongation of *O*-linked fucose influences steps that occur before the proteolytic cleavages that release the Notch intracellular domain¹ (Fig. 6c). Notably, four of the five conserved *O*-linked fucose sites are in or near EGF repeats to which the *Abruptex* class of mutations in *Drosophila* Notch map (N^{Ax}) (Fig. 6b)³⁵. These mutations result in a ligand-dependent hyperactivation of Notch, which implies that this region of Notch normally functions to inhibit Notch activation^{35,36}. This inhibition may also involve cell-autonomous interactions between Notch and its ligands³⁷. N^{Ax} mutants also exhibit an altered sensitivity to Fringe^{32,37}. As both *fringe* and N^{Ax} mutations influence the sensitivity of Notch to its ligands, and both affect the same region of the Notch extracellular domain, we suggest that they impinge on a common regulatory process.

In mammals, which have four Notch proteins and three Fringe proteins, differential substrate specificities may expand the possibilities for Notch modulation by *O*-linked fucose elongation, and mammalian Notch receptors differ with respect to the distribution of their *O*-linked fucose consensus sites^{13,22}. The importance of *O*-linked fucose for mammalian development is evident from the developmental defects in humans with hereditary mutations that reduce GDP-fucose synthesis or transport into the Golgi^{38,39}. Reduced *O*-linked fucose levels result in decreased Notch1 activation (Fig. 4b), but the presence of *O*-linked fucose may not be essential for Notch function. In addition, there are many contexts in which Notch functions normally without *fringe*³. Thus, the elongation of *O*-linked fucose must be viewed as a modulatory rather than an obligatory step in Notch activation. Indeed, there are no identified *fringe* homologues in *Caenorhabditis elegans*, and the Notch-related receptor *lin-12* lacks consensus sites for the addition of *O*-linked fucose.

Recent studies have identified several examples in which heparan sulphate proteoglycans have important roles in signal transduction pathways⁴⁰. Significantly, the direct action of Fringe on Notch represents an unprecedented example of the modulation of ligand–receptor signalling by differential receptor glycosylation. Our demonstration of Fringe-dependent glycosylation of Notch thus provides both a direction for further analysis of the mechanisms that regulate Notch activation, and a new paradigm for the role of glycosylation in development and disease. □

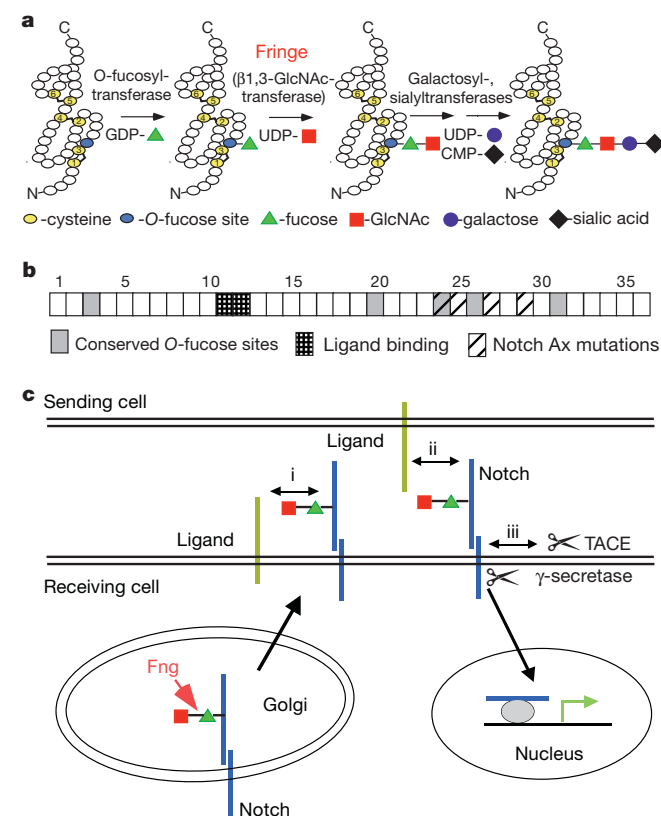


Figure 6 Fringe modulates Notch signalling by elongation of *O*-linked fucose on Notch EGF repeats. **a**, Fringe catalyses elongation of *O*-linked fucose (triangle) on EGF repeats by the addition of a β 1,3-GlcNAc (square). In CHO cells, subsequent steps are addition of a β 1,4-galactose (circle) followed by addition of an α 2,3-sialic acid (diamond). Extent of elongation in *Drosophila* is not yet known. The EGF repeat shown is based on Factor VII (ref. 19). **b**, Significant features within the 36 tandem EGF repeats from Notch: grey boxes, repeats with conserved *O*-linked fucose consensus sites¹³; checked boxes, repeats involved in ligand binding³³; hatched boxes, repeats containing the *Abruptex* mutations³⁵. **c**, Both *O*-linked fucose modification (triangle)¹³ and Fringe-mediated elongation (square) of Notch EGF repeats are predicted to occur in the Golgi where the appropriate transporters for nucleotide sugars (GDP-fucose and UDP-GlcNAc) exist⁴⁶. Fringe-dependent modification may modulate Notch signalling at many steps (indicated by double arrows). For example, cell-autonomous interaction of Notch–DSL ligands (i), intercellular Notch–ligand binding (ii), or the efficacy of Notch–DSL ligand extracellular S2 proteolytic cleavage (iii). The proposed sites of action of the S2 (TACE^{47,48}) and S3 (γ -secretase-like)⁴⁹ proteases are indicated.

Methods

Production of Mfng–AP and Lfng–AP stable CHO transfectants

Wild-type (Pro⁵), Lec1 and Lec13 CHO cells^{16,17,24} were transfected with the pMIRB vector⁴¹ (as vector control), pMIRBMfng–AP (ref. 22) or pMIRBLfng–AP (ref. 22) using Eugene 6 (Roche). Geneticin (Gibco-BRL) resistant clones were characterized for expression level and protein size essentially as described¹⁰.

Analysis of *O*-linked fucose saccharides

The radiolabelling of Lec1 cells, preparation of cell lysate, immunoprecipitation of Notch1 with Na antibody (generated against a conserved intracellular epitope, amino acids 1,789–1,880; ref. 42), release and analysis of *O*-linked fucose saccharides were done as described¹³. The extracellular fragment of mouse Notch1 containing EGF repeats 19–23 was generated by PCR, subcloned into pSecTag2/HygroC (Invitrogen) and stably transfected into Lec1/pMIRB and Lec1/Mfng–AP using Geneporter (Gene Therapy Systems). The EGF 19–23 fragment was purified from medium using Ni²⁺-agarose (Qiagen).

Preparation of Lfng–Fc, Mfng–Fc and ephrin-B2–Fc

HEK 293T cells were transfected with expression plasmids encoding ephrin-B2–Fc (ref. 43), Mfng–Fc or Lfng–Fc (ref. 10) using Eugene 6 (Roche). Fng–Fc proteins were affinity purified from concentrated medium on Protein-A-Sepharose beads (Sigma). The beads were extensively washed and ultimately resuspended in 1 mM Tris–HCl, pH 7, 1 mM MnCl₂, 1 × CPI-EDTA with 20% glycerol and stored at –80 °C. The purity and concentration of bound Fc fusion proteins were determined by densitometric analysis of Coomassie blue stained SDS–PAGE gels against a BSA standard curve.

Expression and purification of *Drosophila* Fringe

Vectors pBSfng:His₆ and pBSfng:myc were generated by cloning an HA tag and six histidines, or a Myc tag into a *Bgl*II site engineered after the last *fng* codon⁴. *fngDEE:HA* and *fngDEE:His₆* were generated by changing codons 237 and 238 from GAT(D) GAC(D) to GAA(E) GAA(E) and subcloning into pUAST or pMthY vectors. Ectopic expression in *Drosophila* was achieved using the UAS–Gal4 system³¹ and *ptc–Gal4* or *fng–Gal4* (gift from J. Botas) drivers. Immunostaining was performed as described⁴.

Expression was induced from either stably or transiently transfected S2 cells with 0.7 mM CuSO₄, and the media collected. After purification on a Ni²⁺-charged column, purity and concentration were determined by densitometric analysis of Coomassie blue stained SDS–PAGE gels.

In vitro GlcNAc-transferase assays

The standard reaction mixture (final volume 50 µl) contained 50 mM HEPES, pH 7.0, 10 mM MnCl₂, 0.1 mg ml^{–1} BSA, and 2 µM UDP-[6-³H]GlcNAc (20 Ci mmol^{–1}; ARC). Other nucleotide sugars (UDP-[6-³H]Gal, UDP-[6-³H]Glc, UDP-[6-³H]GalNAc; ARC) were also 20 Ci mmol^{–1}. The nucleotide sugars were not diluted to ensure that sufficient radioactivity was incorporated for product characterization. Reactions were initiated by the addition of enzyme bound to the affinity adsorbent (Protein-A-Sepharose or Ni-agarose), incubated for 20 min at 37 °C, and terminated by the addition of 450 µl of 50 mM EDTA, pH 8.0. Samples were loaded onto C₁₈ cartridges (100 mg, Waters) equilibrated in water, washed with 10 ml water, and eluted with 1.5 ml of 80% methanol. Product analysis was performed as described^{13,15}.

Expression, purification and labelling of Flag–ECN

Drosophila Flag–ECN was constructed by joining DNA encoding a heterologous signal peptide and Flag epitope from pEF2-SP/Flag (ref. 44), 12 histidine residues and a portion of the *Drosophila* Notch ORF extending from T53 to V1774. Expression of this protein in *Drosophila* under UAS–Gal4 control inhibits Notch signalling, implying that its structure is not affected by the tags⁴⁵ (data not shown). A lysate of S2 cells transiently transfected with Flag–ECN was precleared with Mouse IgG agarose and passed twice through an anti-Flag M2 affinity gel column (Sigma). The column was washed extensively, and beads loaded with Flag–ECN were incubated with S2 cell lysates or purified D-Fng–His in 20 mM HEPES, pH 7.3, 140 mM NaCl, 5 mM KCl, 10 mM MnCl₂ and 0.2% Triton X-100 with 2 µCi of UDP-[³H]GlcNAc (60 Ci mmol^{–1}) and 4 mM ATP for 1 h at 25 °C. Beads were then washed and divided into fractions for scintillation counting, SDS–PAGE/fluorography, and western blotting using mouse anti-Flag M2 (Sigma).

CBF1 luciferase assays

Co-culture assays were performed essentially as described^{21,22}. Duplicate cultures of 2 × 10⁵ CHO, Lec1 or Lec13 cells stably expressing pMIRB vector (control) or pMIRB Mfng–AP or pMIRB Lfng–AP were co-transfected (at a molar ratio of 3:1) with CBF1-luciferase reporter²³ and renilla luciferase plasmids (Promega) using Eugene 6 (Roche). At 12–24 h after transfection, 1 × 10⁶ Jagged1-expressing L cells (J1)²¹ or parental L cells were overlaid. At 24–48 h after transfection, luciferase and renilla luciferase activities were quantitated from cell lysates using a dual luciferase assay (Promega). Jagged1-dependent Notch activation of CBF1 is expressed as a ratio of normalized luciferase from J1 to L cell co-cultures.

Received 24 March; accepted 15 June 2000.

- Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. Notch signalling: Cell fate control and signal integration in development. *Science* **284**, 770–776 (1999).
- Joutel, A. & Tournier-Lasserre, E. Notch signalling pathway and human diseases. *Semin. Cell Dev. Biol.* **9**, 619–625 (1998).
- Panin, V. M. & Irvine, K. D. Modulators of Notch signalling. *Semin. Cell Dev. Biol.* **9**, 609–617 (1998).
- Panin, V. M., Papayannopoulos, V., Wilson, R. & Irvine, K. D. Fringe modulates notch ligand interactions. *Nature* **387**, 908–912 (1997).
- Fleming, R. J., Gu, Y. & Hukriede, N. A. *Serrate*-mediated activation of *Notch* is specifically blocked by the product of the gene *fringe* in the dorsal compartment of the *Drosophila* wing imaginal disc. *Development* **124**, 2973–2981 (1997).
- Klein, T. & Arias, A. M. Interactions among Delta, Serrate and Fringe modulate Notch activity during *Drosophila* wing development. *Development* **125**, 2951–2962 (1998).
- Irvine, K. D. & Wieschaus, E. *fringe*, a boundary-specific signalling molecule, mediates interactions between dorsal and ventral cells during *Drosophila* wing development. *Cell* **79**, 595–606 (1994).
- Irvine, K. D. Fringe, Notch, and making developmental boundaries. *Curr. Opin. Genet. Dev.* **9**, 434–441 (1999).
- Irvine, K. D. & Vogt, T. F. Dorsal–ventral signalling in limb development. *Curr. Opin. Cell Biol.* **9**, 867–876 (1997).
- Johnston, S. H. *et al.* A family of mammalian Fringe genes implicated in boundary determination and the Notch pathway. *Development* **124**, 2245–2254 (1997).
- Cohen, B. *et al.* Fringe boundaries coincide with Notch-dependent patterning centres in mammals and alter Notch-dependent development in *Drosophila*. *Nature Genet.* **16**, 283–288 (1997).
- Yuan, Y. P., Schultz, J., Mlodzik, M. & Bork, P. Secreted fringe-like signaling molecules may be glycosyltransferases. *Cell* **88**, 9–11 (1997).
- Moloney, D. J. *et al.* Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on Epidermal Growth Factor-like modules. *J. Biol. Chem.* **275**, 9604–9611 (2000).
- Harris, R. J. & Spellman, M. W. O-linked fucose and other post-translational modifications unique to EGF modules. *Glycobiology* **3**, 219–224 (1993).
- Moloney, D. J., Lin, A. I. & Haltiwanger, R. S. The O-linked fucose glycosylation pathway: evidence for

- protein specific elongation of O-linked fucose in Chinese hamster ovary cells. *J. Biol. Chem.* **272**, 19046–19050 (1997).
- Stanley, P., Narasimhan, S., Siminovich, L. & Schachter, H. Chinese hamster ovary cells selected for resistance to the cytotoxicity of phytohemagglutinin are deficient in a UDP-N-acetylglucosamine-glycoprotein N-acetylglucosaminyltransferase activity. *Proc. Natl Acad. Sci. USA* **72**, 3323–3327 (1975).
- Robertson, M. A., Etchison, J. R., Robertson, J. S., Summers, D. F. & Stanley, P. Specific changes in the oligosaccharide moieties of VSV grown in different lectin-resistant CHO cells. *Cell* **13**, 515–526 (1978).
- Moloney, D. J. & Haltiwanger, R. S. The O-linked fucose glycosylation pathway: Identification and characterization of a UDP-glucose: O-fucose β1,3-glucosyltransferase. *Glycobiology* **9**, 679–687 (1999).
- Kao, Y. -H. *et al.* The effect of O-fucosylation on the first EGF-like domain from human blood coagulation Factor VII. *Biochemistry* **38**, 7097–7110 (1999).
- Beyer, T. A., Sadler, J. E., Rearick, J. I., Paulson, J. C. & Hill, R. L. Glycosyltransferases and their use in assessing oligosaccharide structure and structure-function relationships. *Adv. Enzymol.* **52**, 23–175 (1981).
- Lindsell, C. E., Shawber, C. J., Boulter, J. & Weinmaster, G. Jagged: A mammalian ligand that activates Notch1. *Cell* **80**, 909–917 (1995).
- Hicks, C., Johnston, S. H., DiSibio, G., Collazo, A., Vogt, T. F. & Weinmaster, G. Fringe differentially modulates Jagged1 and Delta1 signalling through Notch1 and Notch2 receptors in mammalian cells. *Nature Cell Biol.* (in the press).
- Hsieh, J. J., Henkel, T., Salmon, P., Robey, E., Peterson, M. G. & Hayward, S. D. Truncated mammalian Notch1 activates CBF1/RBPJk-repressed genes by a mechanism resembling that of Epstein–Barr virus EBNA2. *Mol. Cell. Biol.* **16**, 952–959 (1996).
- Ripka, J. & Stanley, P. Lectin-resistant CHO cells: Selection of four new pea lectin-resistant phenotypes. *Som. Cell Mol. Gen.* **12**, 51–62 (1986).
- Ripka, J., Adamany, A. & Stanley, P. Two Chinese hamster ovary glycosylation mutants affected in the conversion of GDP-mannose to GDP-fucose. *Arch. Biochem. Biophys.* **249**, 533–545 (1986).
- Sullivan, F. X. *et al.* Molecular cloning of human GDP-mannose 4,6-dehydratase and reconstitution of GDP-fucose biosynthesis in vitro. *J. Biol. Chem.* **273**, 8193–8202 (1998).
- Ohyama, C. *et al.* Molecular cloning and expression of GDP-D-mannose-4,6-dehydratase, a key enzyme for fucose metabolism defective in Lec13 cells. *J. Biol. Chem.* **273**, 14582–14587 (1998).
- Wiggins, C. A. R. & Munro, S. Activity of the yeast *MNN1* α-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc. Natl Acad. Sci. USA* **95**, 7945–7950 (1998).
- Zhang, Y., Malinovsky, V. A., Fiedler, T. J. & Brew, K. Role of a conserved acidic cluster in bovine β1,4 galactosyltransferase-1 probed by mutagenesis of a bacterially expressed recombinant enzyme. *Glycobiology* **9**, 815–822 (1999).
- Kim, J., Irvine, K. D. & Carroll, S. B. Cell recognition, signal induction, and symmetrical gene activation at the dorsal–ventral boundary of the developing *Drosophila* wing. *Cell* **82**, 795–802 (1995).
- Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
- Ju, B. -G. *et al.* Fringe forms a complex with Notch. *Nature* **405**, 191–195 (2000).
- Rebay, I. *et al.* Specific EGF repeats of Notch mediate interactions with Delta and Serrate: Implications for Notch as a multifunctional receptor. *Cell* **67**, 687–699 (1991).
- Klug, K. M. & Muskavitch, M. A. Ligand–receptor interactions and trans-endocytosis of Delta, Serrate and Notch: members of the Notch signaling pathway in *Drosophila*. *J. Cell Sci.* **112**, 3289–3297 (1999).
- Kelley, M. R., Kidd, S., Deutsch, W. A. & Young, M. W. Mutations altering the structure of Epidermal Growth Factor-like coding sequences of the *Drosophila* notch locus. *Cell* **51**, 539–548 (1987).
- De Celis, J. F. & Garcia-Bellido, A. Roles of the *Notch* gene in *Drosophila* wing morphogenesis. *Mech. Dev.* **46**, 109–122 (1994).
- De Celis, J. F. & Bray, S. J. The Abruptex domain of Notch regulates negative interactions between Notch, its ligands and Fringe. *Development* **127**, 1291–1302 (2000).
- Kornfeld, S. Diseases of abnormal protein glycosylation: An emerging area. *J. Clin. Invest.* **101**, 1293–1295 (1998).
- Lubke, T., Marquardt, T., von Figura, K. & Korner, C. A new type of carbohydrate-deficient glycoprotein syndrome due to a decreased import of GDP-fucose into the Golgi. *J. Biol. Chem.* **274**, 25986–25989 (1999).
- Perrimon, N. & Bernfield, M. Specificities of heparan sulphate proteoglycans in developmental processes. *Nature* **404**, 725–728 (2000).
- MacArthur, C. A., Lawshe, A., Shankar, D. B., Heikinheimo, M. & Shackleford, G. M. FGF-8 isoforms differ in NIH3T3 cell transforming potential. *Cell Growth Differ.* **6**, 817–825 (1995).
- Lu, F. M. & Lux, S. E. Constitutively active human Notch1 binds to the transcription factor CBF1 and stimulates transcription through a promoter containing a CBF1-responsive element. *Proc. Natl Acad. Sci. USA* **93**, 5663–5667 (1996).
- Cheng, H. J., Nakamoto, M., Bergemann, A. D. & Flanagan, J. G. Complementary gradients in expression and binding of Elf-1 and Mek-4 in development of the topographic retinotectal projection map. *Cell* **82**, 371–381 (1995).
- Kotenko, S. V., Sacconi, S., Izotova, L. S., Mironchitschenko, O. V. & Pestka, S. Human cytomegalovirus harbors its own unique interleukin-10 homolog (cmvIL-10). *Proc. Natl Acad. Sci. USA* **97**, 1695–1700 (2000).
- Rebay, I., Fehon, R. G. & Artavanis-Tsakonas, S. Specific truncations of *Drosophila* Notch define dominant activated and dominant negative forms of the receptor. *Cell* **74**, 319–329 (1993).
- Hirschberg, C. B., Robbins, P. W. & Abejón, C. Transporters of nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus. *Annu. Rev. Biochem.* **67**, 49–69 (1998).
- Brou, C. *et al.* A novel proteolytic cleavage involved in Notch signalling: The role of the disintegrin-metalloprotease TACE. *Mol. Cell* **5**, 207–216 (2000).
- Mumm, J. S. *et al.* A ligand-induced extracellular cleavage regulates γ-secretase-like proteolytic activation of Notch1. *Mol. Cell* **5**, 197–206 (2000).
- De Strooper, B. *et al.* A presenilin-1-dependent γ-secretase-like protease mediates release of Notch intracellular fragment. *Nature* **398**, 518–522 (1999).



Acknowledgements

We thank K. Matta for disaccharide standards; K. Severinov for advice on protein purification; T. Correia, S. Narula and S. Sundaram for technical assistance; the Developmental Studies Hybridoma Bank; S. Lux and S. Kotenko for antibodies; G. Weinmaster for Jagged1-expressing mouse L cells and co-culture reagents; J. Flanagan for the AP-Tag plasmid; D. Hayward for CBF1-luciferase reporter plasmid JH26; J. Nye for the mouse Notch1 plasmid; S. Kotenko for Flag expression plasmid; and J. Botas and the Bloomington Stock Center for *Drosophila* stocks. We thank our colleagues for comments on the manuscript. S.H.J. was supported as a Harold W. Dodds Fellow of the Princeton

University Graduate Program. Research in K.D.I.'s laboratory was supported by a GM NIH grant, and by Charles and Joanna Busch fellowships to V.P. and R.W. Research in R.S.H.'s laboratory was supported by a Grant Award from Neose Technologies, Inc. Research in P.S.'s laboratory was supported by a NIH/NCI grant and a NCI Comprehensive Cancer Center grant. Research in T.F.V.'s lab was supported by an NIH grant and a gift to the Department of Molecular Biology by the Rathmann Family Foundation.

Correspondence and requests for materials should be addressed to R.S.H. (e-mail: rhaltiwanger@ms.cc.sunysb.edu), K.D.I. (e-mail: irvine@waksman.rutgers.edu), P. S. (e-mail: stanley_@aecom.yu.edu) or T.F.V. (e-mail: thomas_vogt@merck.com).

Characterizing the nonlinear growth of large-scale structure in the Universe

Peter Coles* & Lung-Yih Chiang†

* School of Physics & Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK

† Astronomy Unit, School of Mathematical Sciences, Queen Mary & Westfield College, University of London, London E1 4NS, UK

The local Universe displays a rich hierarchical pattern of galaxy clusters and superclusters^{1,2}. The early Universe, however, was almost smooth, with only slight ‘ripples’, as seen in the cosmic microwave background radiation³. Models of the evolution of cosmic structure link these observations through the effect of gravity, because the small initially overdense fluctuations are predicted to attract additional mass as the Universe expands⁴. During the early stages of this expansion, the ripples evolve independently, like linear waves on the surface of deep water. As the structures grow in mass, they interact with each other in nonlinear ways, more like waves breaking in shallow water. We have recently shown⁵ how cosmic structure can be characterized by phase correlations associated with these nonlinear interactions, but it was not clear how to use that information to obtain quantitative insights into the growth of structures. Here we report a method of revealing phase information, and show quantitatively how this relates to the formation of filaments, sheets and clusters of galaxies by nonlinear collapse. We develop a statistical method based on information entropy to separate linear from nonlinear effects, and thereby are able to disentangle those aspects of galaxy clustering that arise from initial conditions (the ripples) from the subsequent dynamical evolution.

In most popular versions of the “gravitational instability” model for the origin of cosmic structure, particularly those involving

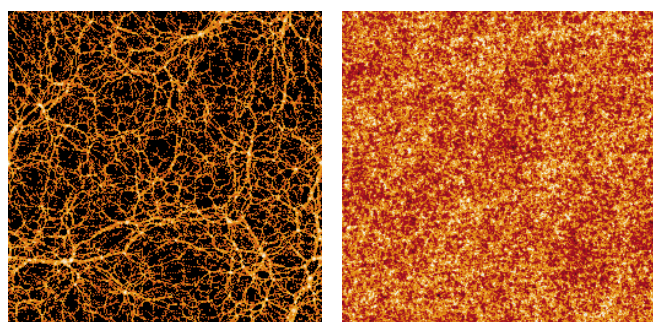


Figure 1 The importance of phase coupling. Left, numerical simulation of galaxy clustering; right, a version of the same picture generated by randomly reshuffling the phases between Fourier modes of the original picture. Because the amplitude of each Fourier mode is unchanged in this operation, these two pictures have exactly the same power-spectrum, $P(k) \propto |\tilde{\delta}(\mathbf{k})|^2$, but have totally different morphology. The shortcomings of $P(k)$ can be partly ameliorated by defining higher-order quantities such as the bispectrum^{20–22} or correlations²³ of $\tilde{\delta}(\mathbf{k})^2$. The bispectrum and higher-order polyspectra vanish for gaussian fields, but in a non-gaussian field they may be non-zero. The usefulness of these and related quantities therefore lies in the fact that they encode some information about non-linearity and non-gaussianity. The bispectrum, for example, measures the phase coupling induced by quadratic nonlinearities, and so on for higher orders. However, an infinite hierarchy of such moments is necessary to specify the properties of a general random field in a statistical sense. But because not every distribution is specified by its moments, this need not be complete.

cosmic inflation⁶, the initial fluctuations that seeded the structure-formation process form a gaussian random field⁷. Deviations from uniformity are expressed in terms of the density contrast $\delta(\mathbf{x})$, defined by

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \rho_0}{\rho_0} \quad (1)$$

where ρ_0 is the average density and $\rho(\mathbf{x})$ is the local matter density. Because the initial perturbations evolve linearly, it is useful to expand $\rho(\mathbf{x})$ as a Fourier superposition of plane waves:

$$\delta(\mathbf{x}) = \sum \tilde{\delta}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x}) \quad (2)$$

The Fourier transform $\tilde{\delta}(\mathbf{k})$ is complex, and therefore possesses both amplitude $|\tilde{\delta}(\mathbf{k})|$ and phase ϕ_k :

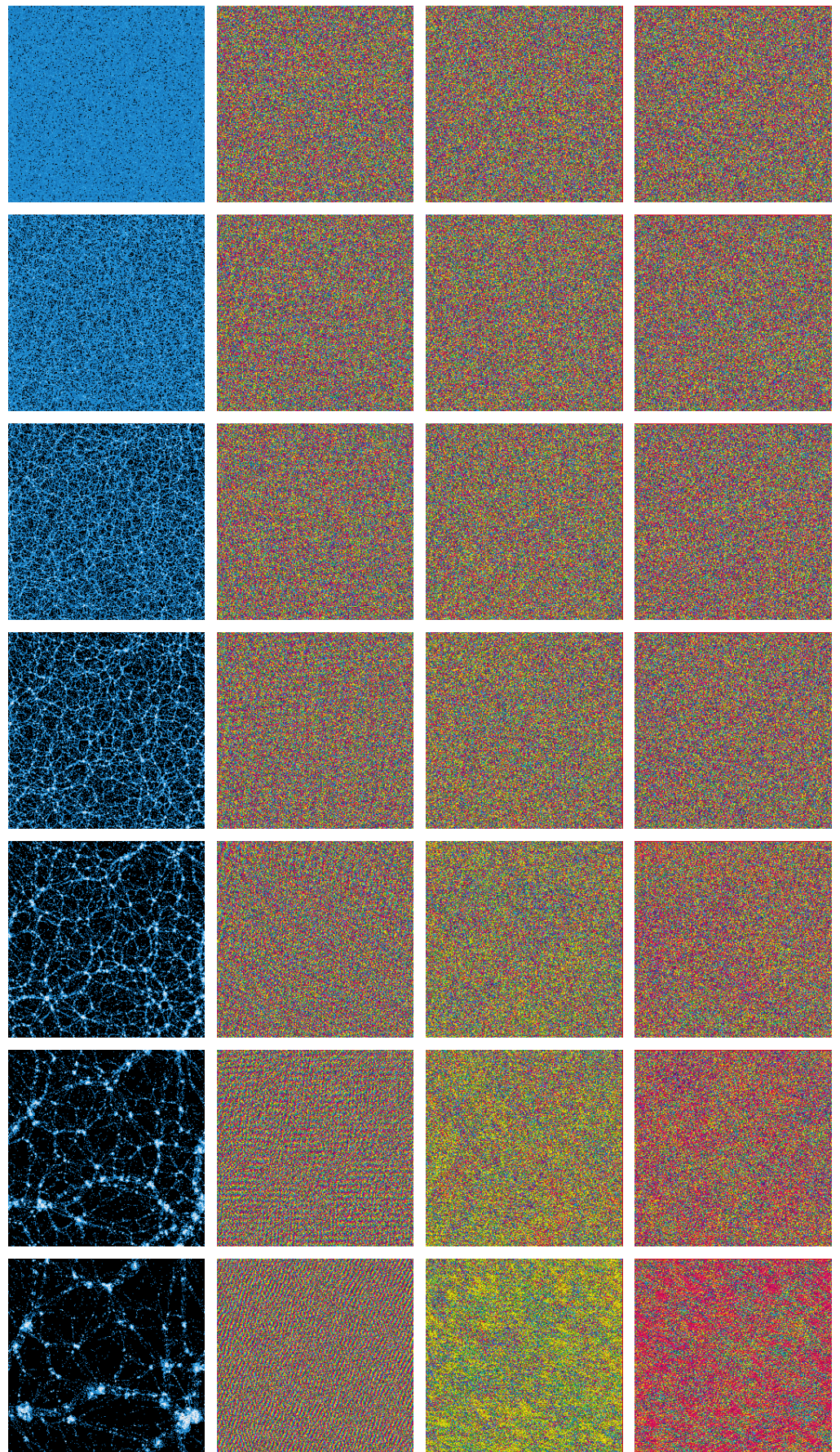
$$\tilde{\delta}(\mathbf{k}) = |\tilde{\delta}(\mathbf{k})| \exp(i\phi_k) \quad (3)$$

Gaussian random fields possess Fourier modes whose real and imaginary parts are independently distributed. In other words, they have phase angles ϕ_k that are independently distributed and uniformly random on the interval $[0, 2\pi]$. When fluctuations are small—that is, during the linear regime—the Fourier modes evolve independently and their phases remain random. In the later stages of evolution, however, wave modes begin to couple together⁴. In this regime, the phases become non-random and the density field becomes highly non-gaussian. Phase coupling is therefore an



Figure 2 The representation of colour hue on a circle. In colour image display devices, each pixel represents the intensity and colour at that position in the image^{14,24}. The quantitative specification of colour involves three coordinates describing the location of that pixel in an abstract colour space, designed to reflect as accurately as possible the eye’s response to light of different wavelengths. In many devices, this colour space is defined in terms of the amount of red, green or blue required to construct the appropriate tone; hence the RGB colour scheme. The scheme we are particularly interested in, the HSB scheme, is based on three different parameters: hue, saturation and brightness. Hue is the term used to distinguish between different basic colours (blue, yellow, red and so on). Saturation refers to the purity of the colour, defined by how much white is mixed with it. A saturated red hue would be a very bright red, whereas a less saturated red would be pink. Brightness indicates the overall intensity of the pixel on a grey scale. The HSB colour model is particularly useful because of the properties of the ‘hue’ parameter, which is defined as a circular variable. On the colour circle shown, the primary hues (red, green and blue) are 120° apart from one another (at 0°, 120° and 240° respectively), while the complementary tones yellow, cyan and magenta are at 60°, 180° and 300°. Red, of course, also appears at 360° and so on, so this parameter is truly circular in the same way as phases are. Our visualization method uses the hue parameter to encode the Fourier phases.

Figure 3 The evolution of phase coupling. This figure shows a sequence of snapshots from a two-dimensional N -body simulation^{25,26} starting with gaussian initial conditions. The initial power-spectrum was a power-law, $P(k) \propto k^n$, with $n = 0$. The left-hand column shows the development of hierarchical clustering through the emergence in the density field of the characteristic network of voids and filaments that typifies the action of gravitational instability. The next column shows the colour-coded phases, and the third and fourth columns show phase gradients in the x -direction and y -direction, respectively. Each row corresponds to a particular timestep, with time progressing downwards. The early stages display no phase structure, but patterns emerge as time increases. The phase images develop a series of bands and stripes, first appearing as large-scale features in reciprocal space. As the simulation evolves, the characteristic k -space scale of these features gets smaller. This behaviour is related to the increasing characteristic scale of features in the density maps in the first column: larger features in real space correspond to smaller features in the reciprocal space. Phase correlations between k -modes reveal themselves as structure in the distribution of the D_{ik} , chiefly through a domination of some hues over others. Starting from the initial data, in which they are distributed evenly, we see the gradual appearance of regions where certain hues dominate. By the final stage, the third column is largely green and the last column is largely red, indicating strongly coupled phases in both directions. Animations of this effect can be viewed at our website, <http://www.nottingham.ac.uk/~ppzpc/phases/index.html>.



important consequence of nonlinear gravitational processes if the initial conditions are gaussian, and is a potentially powerful signature to exploit in statistical tests of this class of models; see Fig. 1.

The information needed to fully specify a non-gaussian field (or, in a wider context, the information needed to define an image⁸) resides in the complete set of Fourier phases. Unfortunately, relatively little is known about the behaviour of Fourier phases in the nonlinear regime of gravitational clustering^{9–14}, but it is essential to understand phase correlations in order to design efficient statistical tools for the analysis of clustering data. A first step on the road to a useful quantitative description of phase information is to represent it visually. We do this using colour, as shown in Fig. 2. To view the phase coupling in an N -body simulation, we Fourier-transform the density field; this produces a complex array containing the real (R) and imaginary (I) parts of the transformed 'image', with the pixels in this array labelled by wavenumber $\mathbf{k}$ rather than position $\mathbf{x}$. The phase for each wavenumber, given by $\phi = \arctan(I/R)$, is then represented as a hue for that pixel.

The rich pattern of phase information revealed by this method (see Fig. 3) can be quantified, and related to the gravitational dynamics of its origin. For example, in our analysis of phase coupling⁵ we introduced a quantity D_k :

$$D_k \equiv \phi_{k+1} - \phi_k \quad (4)$$

This quantity measures the difference in phase of modes with neighbouring wavenumbers in one dimension. We refer to D_k as the phase gradient. To apply this idea to a two-dimensional simulation, we simply calculate gradients in the x and y directions independently. Because the difference between two circular random variables is itself a circular random variable, the distribution of D_k should initially be uniform. As the fluctuations evolve waves begin to collapse, spawning higher-frequency modes in phase with the original¹⁵. These then interact with other waves to produce the non-uniform distribution of D_k seen in Fig. 3.

It is necessary to develop quantitative measures of phase information that can describe the structure displayed in the colour representations. In the beginning, the phases ϕ_k are random and so are the D_k obtained from them. This corresponds to a state of minimal information, or in other words, maximum entropy. As information flows into the phases, the information content must increase and the entropy decrease. This can be quantified by defining an information entropy for the set of phase gradients⁵. We construct a frequency distribution, $f(D)$, of the values of D_k obtained from the whole map. The entropy is then defined as

$$S(D) = - \int f(D) \log[f(D)] dD \quad (5)$$

where the integral is taken over all values of D , that is, from 0 to 2π . The use of D , rather than ϕ itself, to define entropy is one way of accounting for the lack of translation invariance of ϕ , a problem that was missed in previous attempts to quantify phase entropy¹⁶. A uniform distribution of D is a state of maximum entropy (minimum information), corresponding to gaussian initial conditions (random phases). This maximal value of $S_{\max} = \log(2\pi)$ is a characteristic of gaussian fields. As the system evolves, it moves into states of greater information content (that is, lower entropy). The scaling of S with clustering growth displays interesting properties⁵, establishing an important link between the spatial pattern and the physical processes driving clustering growth. This phase information is a unique 'fingerprint' of gravitational instability, and it therefore also furnishes statistical tests of the presence of any initial non-gaussianity^{17–19}. □

Received 17 January; accepted 19 May 2000.

1. Saunders, W. *et al.* The density field of the local Universe. *Nature* **349**, 32–38 (1991).
2. Shectman, S. *et al.* The Las Campanas redshift survey. *Astrophys. J.* **470**, 172–188 (1996).
3. Smoot, G. F. *et al.* Structure in the COBE differential microwave radiometer first-year maps. *Astrophys. J.* **396**, L1–L4 (1992).

4. Peebles, P. J. E. *The Large-scale Structure of the Universe* (Princeton Univ. Press, Princeton, 1980).
5. Chiang, L.-Y. & Coles, P. Phase information and the evolution of cosmological density perturbations. *Mon. Not. R. Astron. Soc.* **311**, 809–824 (2000).
6. Guth, A. H. & Pi, S.-Y. Fluctuations in the new inflationary universe. *Phys. Rev. Lett.* **49**, 1110–1113 (1982).
7. Bardeen, J. M., Bond, J. R., Kaiser, N. & Szalay, A. S. The statistics of peaks of Gaussian random fields. *Astrophys. J.* **304**, 15–61 (1986).
8. Oppenheim, A. V. & Lim, J. S. The importance of phase in signals. *Proc. IEEE* **69**, 529–541 (1981).
9. Ryden, B. S. & Gramann, M. Phase shifts in gravitationally evolving density fields. *Astrophys. J.* **383**, L33–L36 (1991).
10. Scherrer, R. J., Melott, A. L. & Shandarin, S. F. A quantitative measure of phase correlations in density fields. *Astrophys. J.* **377**, 29–35 (1991).
11. Soda, J. & Suto, Y. Nonlinear gravitational evolution of phases and amplitudes in one-dimensional cosmological density fields. *Astrophys. J.* **396**, 379–394 (1992).
12. Jain, B. & Bertschinger, E. Self-similar evolution of gravitational clustering: is $n = 1$ special? *Astrophys. J.* **456**, 43–54 (1996).
13. Jain, B. & Bertschinger, E. Self-similar evolution of gravitational clustering: N -body simulations of the $n = -2$ spectrum. *Astrophys. J.* **509**, 517–530 (1998).
14. Thornton, A. L. *Colour Object Recognition Using a Complex Colour Representation and the Frequency Domain*. Thesis, Univ. Reading (1998).
15. Shandarin, S. F. & Zel'dovich, Ya. B. The large-scale structure: turbulence, intermittency, structures in a self-gravitating medium. *Rev. Mod. Phys.* **61**, 185–220 (1989).
16. Polygiannakis, J. M. & Moussas, X. Detection of nonlinear dynamics in solar wind and a comet using phase-correlation measures. *Sol. Phys.* **158**, 159–172 (1995).
17. Ferreira, P. G., Magueijo, J. & Górski, K. M. Evidence for non-Gaussianity in the COBE DMR 4-year sky maps. *Astrophys. J.* **503**, L1–L4 (1998).
18. Pando, J., Valls-Gabaud, D. & Fang, L. Evidence for scale-scale correlations in the cosmic microwave background radiation. *Phys. Rev. Lett.* **81**, 4568–4571 (1998).
19. Bromley, B. C. & Tegmark, M. Is the cosmic microwave background really non-gaussian? *Astrophys. J.* **524**, L79–L82 (1999).
20. Matarrese, S., Verde, L. & Heavens, A. F. Large-scale bias in the universe: bispectrum method. *Mon. Not. R. Astron. Soc.* **290**, 651–662 (1997).
21. Scoccimarro, R., Couchman, H. M. P. & Frieman, J. A. The bispectrum as a signature of gravitational instability in redshift space. *Astrophys. J.* **517**, 531–540 (1999).
22. Verde, L., Wang, L., Heavens, A. F. & Kamionkowski, M. Large-scale structure, the cosmic microwave background, and primordial non-Gaussianity. *Mon. Not. R. Astron. Soc.* **313**, 141–147 (2000).
23. Stirling, A. J. & Peacock, J. A. Power correlations in cosmology: Limits on primordial non-Gaussian density fields. *Mon. Not. R. Astron. Soc.* **283**, L99–L104 (1996).
24. Foley, J. D. & Van Dam, A. *Fundamentals of Interactive Computer Graphics* (Addison-Wesley, Reading, Massachusetts, 1982).
25. Melott, A. L. & Shandarin, S. F. Generation of large-scale cosmological structures by gravitational clustering. *Nature* **346**, 633–635 (1990).
26. Beacom, J. F., Dominik, K. G., Melott, A. L., Perkins, S. P. & Shandarin, S. F. Gravitational clustering in the expanding Universe—controlled high resolution studies in two dimensions. *Astrophys. J.* **372**, 351–363 (1991).

Correspondence and requests for materials should be addressed to P. C. (e-mail: Peter.Coles@Nottingham.ac.uk). Colour animations of phase evolution from a set of N -body experiments, including the one shown in Fig. 3, can be viewed at <http://www.nottingham.ac.uk/~ppzpc/phases/index.html>.

Error and attack tolerance of complex networks

Réka Albert, Hawoong Jeong & Albert-László Barabási

Department of Physics, 225 Nieuwland Science Hall, University of Notre Dame, Notre Dame, Indiana 46556, USA

Many complex systems display a surprising degree of tolerance against errors. For example, relatively simple organisms grow, persist and reproduce despite drastic pharmaceutical or environmental interventions, an error tolerance attributed to the robustness of the underlying metabolic network¹. Complex communication networks² display a surprising degree of robustness: although key components regularly malfunction, local failures rarely lead to the loss of the global information-carrying ability of the network. The stability of these and other complex systems is often attributed to the redundant wiring of the functional web defined by the systems' components. Here we demonstrate that error tolerance is not shared by all redundant systems: it is displayed only by a class of inhomogeneously wired networks,

called scale-free networks, which include the World-Wide Web^{3–5}, the Internet⁶, social networks⁷ and cells⁸. We find that such networks display an unexpected degree of robustness, the ability of their nodes to communicate being unaffected even by unrealistically high failure rates. However, error tolerance comes at a high price in that these networks are extremely vulnerable to attacks (that is, to the selection and removal of a few nodes that play a vital role in maintaining the network's connectivity). Such error tolerance and attack vulnerability are generic properties of communication networks.

The increasing availability of topological data on large networks, aided by the computerization of data acquisition, had led to great advances in our understanding of the generic aspects of network structure and development^{9–16}. The existing empirical and theoretical results indicate that complex networks can be divided into two major classes based on their connectivity distribution $P(k)$, giving the probability that a node in the network is connected to k other nodes. The first class of networks is characterized by a $P(k)$ that peaks at an average $\langle k \rangle$ and decays exponentially for large k . The most investigated examples of such exponential networks are the random graph model of Erdős and Rényi^{9,10} and the small-world model of Watts and Strogatz¹¹, both leading to a fairly homogeneous network, in which each node has approximately the same number of links, $k \approx \langle k \rangle$. In contrast, results on the World-Wide Web (WWW)^{3–5}, the Internet⁶ and other large networks^{17–19} indicate that many systems belong to a class of inhomogeneous networks, called scale-free networks, for which $P(k)$ decays as a power-law, that is $P(k) \sim k^{-\gamma}$, free of a characteristic scale. Whereas the probability that a node has a very large number of connections ($k \gg \langle k \rangle$) is practically prohibited in exponential networks, highly connected nodes are statistically significant in scale-free networks (Fig. 1).

We start by investigating the robustness of the two basic connectivity distribution models, the Erdős–Rényi (ER) model^{9,10} that produces a network with an exponential tail, and the scale-free model¹⁷ with a power-law tail. In the ER model we first define the N nodes, and then connect each pair of nodes with probability p . This algorithm generates a homogeneous network (Fig. 1), whose connectivity follows a Poisson distribution peaked at $\langle k \rangle$ and decaying exponentially for $k \gg \langle k \rangle$.

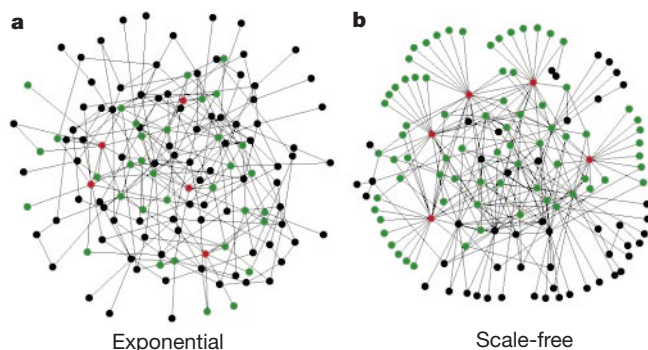


Figure 1 Visual illustration of the difference between an exponential and a scale-free network. **a**, The exponential network is homogeneous: most nodes have approximately the same number of links. **b**, The scale-free network is inhomogeneous: the majority of the nodes have one or two links but a few nodes have a large number of links, guaranteeing that the system is fully connected. Red, the five nodes with the highest number of links; green, their first neighbours. Although in the exponential network only 27% of the nodes are reached by the five most connected nodes, in the scale-free network more than 60% are reached, demonstrating the importance of the connected nodes in the scale-free network. Both networks contain 130 nodes and 215 links ($\langle k \rangle = 3.3$). The network visualization was done using the Pajek program for large network analysis: (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/pajekman.htm>).

The inhomogeneous connectivity distribution of many real networks is reproduced by the scale-free model^{17,18} that incorporates two ingredients common to real networks: growth and preferential attachment. The model starts with m_0 nodes. At every time step t a new node is introduced, which is connected to m of the already-existing nodes. The probability Π_i that the new node is connected to node i depends on the connectivity k_i of node i such that $\Pi_i = k_i / \sum_j k_j$. For large t the connectivity distribution is a power-law following $P(k) = 2m^2/k^3$.

The interconnectedness of a network is described by its diameter d , defined as the average length of the shortest paths between any two nodes in the network. The diameter characterizes the ability of two nodes to communicate with each other: the smaller d is, the shorter is the expected path between them. Networks with a very large number of nodes can have quite a small diameter; for example, the diameter of the WWW, with over 800 million nodes²⁰, is around 19 (ref. 3), whereas social networks with over six billion individuals

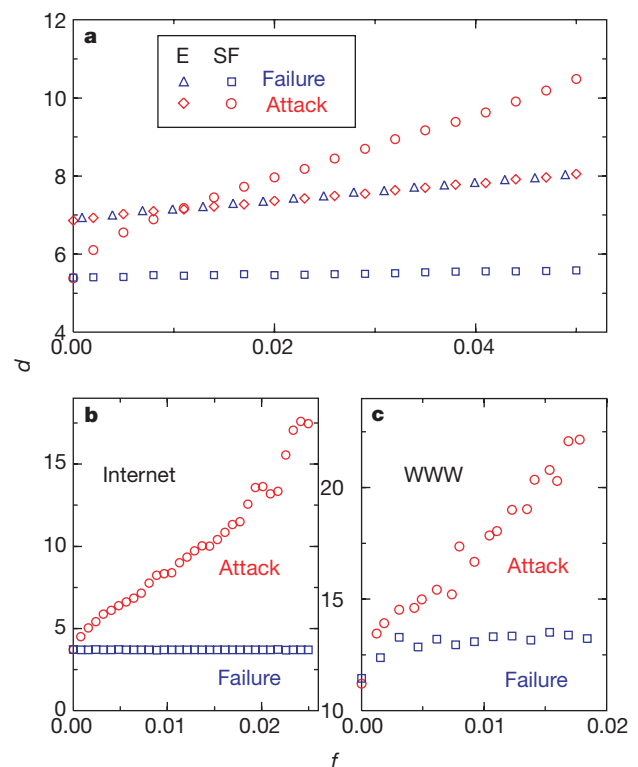


Figure 2 Changes in the diameter d of the network as a function of the fraction f of the removed nodes. **a**, Comparison between the exponential (E) and scale-free (SF) network models, each containing $N = 10,000$ nodes and 20,000 links (that is, $\langle k \rangle = 4$). The blue symbols correspond to the diameter of the exponential (triangles) and the scale-free (squares) networks when a fraction f of the nodes are removed randomly (error tolerance). Red symbols show the response of the exponential (diamonds) and the scale-free (circles) networks to attacks, when the most connected nodes are removed. We determined the f dependence of the diameter for different system sizes ($N = 1,000; 5,000; 20,000$) and found that the obtained curves, apart from a logarithmic size correction, overlap with those shown in **a**, indicating that the results are independent of the size of the system. We note that the diameter of the unperturbed ($f = 0$) scale-free network is smaller than that of the exponential network, indicating that scale-free networks use the links available to them more efficiently, generating a more interconnected web. **b**, The changes in the diameter of the Internet under random failures (squares) or attacks (circles). We used the topological map of the Internet, containing 6,209 nodes and 12,200 links ($\langle k \rangle = 3.4$), collected by the National Laboratory for Applied Network Research (<http://moat.nlanr.net/Routing/rawdata/>). **c**, Error (squares) and attack (circles) survivability of the World-Wide Web, measured on a sample containing 325,729 nodes and 1,498,353 links³, such that $\langle k \rangle = 4.59$.

are believed to have a diameter of around six²¹. To compare the two network models properly, we generated networks that have the same number of nodes and links, such that $P(k)$ follows a Poisson distribution for the exponential network, and a power law for the scale-free network.

To address the error tolerance of the networks, we study the changes in diameter when a small fraction f of the nodes is removed. The malfunctioning (absence) of any node in general increases the distance between the remaining nodes, as it can eliminate some paths that contribute to the system's interconnectedness. Indeed, for the exponential network the diameter increases monotonically with f (Fig. 2a); thus, despite its redundant wiring (Fig. 1), it is increasingly difficult for the remaining nodes to communicate with each other. This behaviour is rooted in the homogeneity of the network: since all nodes have approximately the same number of links, they all contribute equally to the network's diameter, thus the removal of each node causes the same amount of damage. In contrast, we observe a drastically different and surprising behaviour for the scale-free network (Fig. 2a): the diameter remains unchanged under an increasing level of errors. Thus even when as many as 5% of

the nodes fail, the communication between the remaining nodes in the network is unaffected. This robustness of scale-free networks is rooted in their extremely inhomogeneous connectivity distribution: because the power-law distribution implies that the majority of nodes have only a few links, nodes with small connectivity will be selected with much higher probability. The removal of these 'small' nodes does not alter the path structure of the remaining nodes, and thus has no impact on the overall network topology.

An informed agent that attempts to deliberately damage a network will not eliminate the nodes randomly, but will preferentially target the most connected nodes. To simulate an attack we first remove the most connected node, and continue selecting and removing nodes in decreasing order of their connectivity k . Measuring the diameter of an exponential network under attack, we find that, owing to the homogeneity of the network, there is no substantial difference whether the nodes are selected randomly or in decreasing order of connectivity (Fig. 2a). On the other hand, a drastically different behaviour is observed for scale-free networks. When the most connected nodes are eliminated, the diameter of the scale-free network increases rapidly, doubling its original value if 5% of the nodes are removed. This vulnerability to attacks is rooted in the inhomogeneity of the connectivity distribution: the connectivity is maintained by a few highly connected nodes (Fig. 1b), whose removal drastically alters the network's topology, and

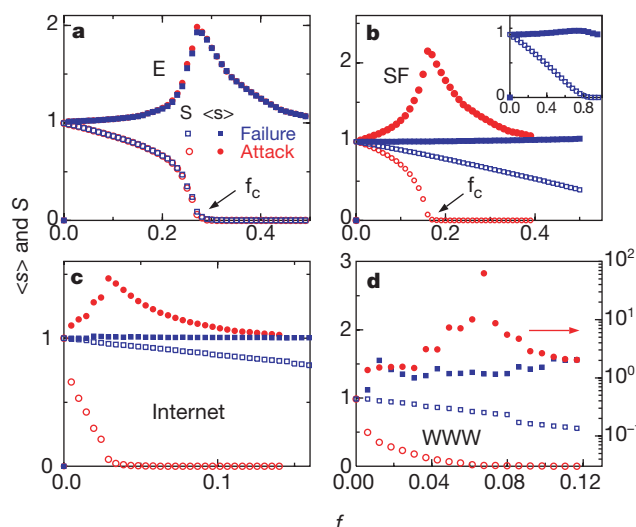


Figure 3 Network fragmentation under random failures and attacks. The relative size of the largest cluster S (open symbols) and the average size of the isolated clusters $\langle s \rangle$ (filled symbols) as a function of the fraction of removed nodes f for the same systems as in Fig. 2. The size S is defined as the fraction of nodes contained in the largest cluster (that is, $S = 1$ for $f = 0$). **a**, Fragmentation of the exponential network under random failures (squares) and attacks (circles). **b**, Fragmentation of the scale-free network under random failures (blue squares) and attacks (red circles). The inset shows the error tolerance curves for the whole range of f , indicating that the main cluster falls apart only after it has been completely deflated. We note that the behaviour of the scale-free network under errors is consistent with an extremely delayed percolation transition: at unrealistically high error rates ($f_{\max} \approx 0.75$) we do observe a very small peak in $\langle s \rangle$ ($\langle s_{\max} \rangle \approx 1.06$) even in the case of random failures, indicating the existence of a critical point. For **a** and **b** we repeated the analysis for systems of sizes $N = 1,000, 5,000$ and $20,000$, finding that the obtained S and $\langle s \rangle$ curves overlap with the one shown here, indicating that the overall clustering scenario and the value of the critical point is independent of the size of the system. **c**, **d**, Fragmentation of the Internet (**c**) and WWW (**d**), using the topological data described in Fig. 2. The symbols are the same as in **b**. $\langle s \rangle$ in **d** in the case of attack is shown on a different scale, drawn in the right side of the frame. Whereas for small f we have $\langle s \rangle \approx 1.5$, at $f_c^w = 0.067$ the average fragment size abruptly increases, peaking at $\langle s_{\max} \rangle \approx 60$, then decays rapidly. For the attack curve in **d** we ordered the nodes as a function of the number of outgoing links, k_{out} . We note that while the three studied networks, the scale-free model, the Internet and the WWW have different γ , $\langle k \rangle$ and clustering coefficient¹¹, their response to attacks and errors is identical. Indeed, we find that the difference between these quantities changes only f_c and the magnitude of d , S and $\langle s \rangle$, but not the nature of the response of these networks to perturbations.

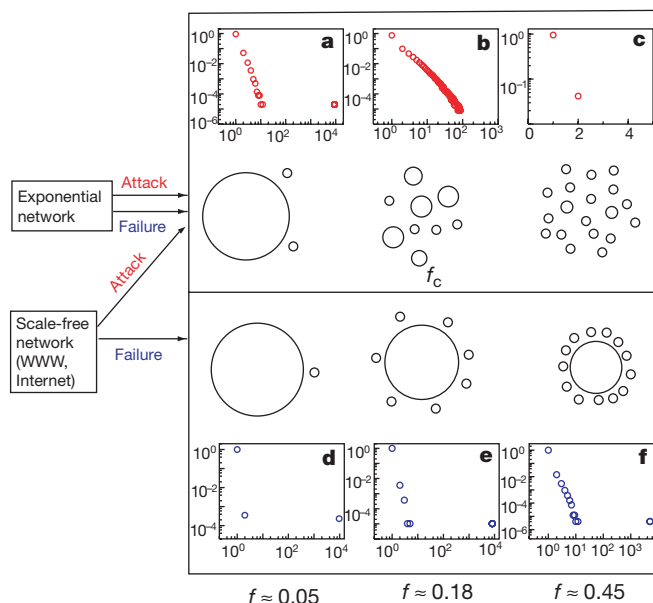


Figure 4 Summary of the response of a network to failures or attacks. **a–f**, The cluster size distribution for various values of f when a scale-free network of parameters given in Fig. 3b is subject to random failures (**a–c**) or attacks (**d–f**). Upper panels, exponential networks under random failures and attacks and scale-free networks under attacks behave similarly. For small f , clusters of different sizes break down, although there is still a large cluster. This is supported by the cluster size distribution: although we see a few fragments of sizes between 1 and 16, there is a large cluster of size 9,000 (the size of the original system being 10,000). At a critical f_c (see Fig. 3) the network breaks into small fragments between sizes 1 and 100 (**b**) and the large cluster disappears. At even higher f (**c**) the clusters are further fragmented into single nodes or clusters of size two. Lower panels, scale-free networks follow a different scenario under random failures: the size of the largest cluster decreases slowly as first single nodes, then small clusters break off. Indeed, at $f = 0.05$ only single and double nodes break off (**d**). At $f = 0.18$, the network is fragmented (**b**) under attack, but under failures the large cluster of size 8,000 coexists with isolated clusters of sizes 1 to 5 (**e**). Even for an unrealistically high error rate of $f = 0.45$ the large cluster persists, the size of the broken-off fragments not exceeding 11 (**f**).

decreases the ability of the remaining nodes to communicate with each other.

When nodes are removed from a network, clusters of nodes whose links to the system disappear may be cut off (fragmented) from the main cluster. To better understand the impact of failures and attacks on the network structure, we next investigate this fragmentation process. We measure the size of the largest cluster, S , shown as a fraction of the total system size, when a fraction f of the nodes are removed either randomly or in an attack mode. We find that for the exponential network, as we increase f , S displays a threshold-like behaviour such that for $f > f_c^e \approx 0.28$ we have $S \approx 0$. Similar behaviour is observed when we monitor the average size $\langle s \rangle$ of the isolated clusters (that is, all the clusters except the largest one), finding that $\langle s \rangle$ increases rapidly until $\langle s \rangle \approx 2$ at f_c^e , after which it decreases to $\langle s \rangle = 1$. These results indicate the following breakdown scenario (Fig. 3a). For small f , only single nodes break apart, $\langle s \rangle \approx 1$, but as f increases, the size of the fragments that fall off the main cluster increases, displaying unusual behaviour at f_c^e . At f_c^e the system falls apart; the main cluster breaks into small pieces, leading to $S \approx 0$, and the size of the fragments, $\langle s \rangle$, peaks. As we continue to remove nodes ($f > f_c^e$), we fragment these isolated clusters, leading to a decreasing $\langle s \rangle$. Because the ER model is equivalent to infinite dimensional percolation²², the observed threshold behaviour is qualitatively similar to the percolation critical point.

However, the response of a scale-free network to attacks and failures is rather different (Fig. 3b). For random failures no threshold for fragmentation is observed; instead, the size of the largest cluster slowly decreases. The fact that $\langle s \rangle \approx 1$ for most f values indicates that the network is deflated by nodes breaking off one by one, the increasing error level leading to the isolation of single nodes only, not clusters of nodes. Thus, in contrast with the catastrophic fragmentation of the exponential network at f_c^e , the scale-free network stays together as a large cluster for very high values of f , providing additional evidence of the topological stability of these networks under random failures. This behaviour is consistent with the existence of an extremely delayed critical point (Fig. 3) where the network falls apart only after the main cluster has been completely deflated. On the other hand, the response to attack of the scale-free network is similar (but swifter) to the response to attack and failure of the exponential network (Fig. 3b): at a critical threshold $f_c^f \approx 0.18$, smaller than the value $f_c^e \approx 0.28$ observed for the exponential network, the system breaks apart, forming many isolated clusters (Fig. 4).

Although great efforts are being made to design error-tolerant and low-yield components for communication systems, little is known about the effect of errors and attacks on the large-scale connectivity of the network. Next, we investigate the error and attack tolerance of two networks of increasing economic and strategic importance: the Internet and the WWW.

Faloutsos *et al.*⁶ investigated the topological properties of the Internet at the router and inter-domain level, finding that the connectivity distribution follows a power-law, $P(k) \sim k^{-2.48}$. Consequently, we expect that it should display the error tolerance and attack vulnerability predicted by our study. To test this, we used the latest survey of the Internet topology, giving the network at the inter-domain (autonomous system) level. Indeed, we find that the diameter of the Internet is unaffected by the random removal of as high as 2.5% of the nodes (an order of magnitude larger than the failure rate (0.33%) of the Internet routers²³), whereas if the same percentage of the most connected nodes are eliminated (attack), d more than triples (Fig. 2b). Similarly, the large connected cluster persists for high rates of random node removal, but if nodes are removed in the attack mode, the size of the fragments that break off increases rapidly, the critical point appearing at $f_c^f \approx 0.03$ (Fig. 3b).

The WWW forms a huge directed graph whose nodes are documents and edges are the URL hyperlinks that point from one

document to another, its topology determining the search engines' ability to locate information on it. The WWW is also a scale-free network: the probabilities $P_{\text{out}}(k)$ and $P_{\text{in}}(k)$ that a document has k outgoing and incoming links follow a power-law over several orders of magnitude, that is, $P(k) \sim k^{-\gamma}$, with $\gamma_{\text{in}} = 2.1$ and $\gamma_{\text{out}} = 2.45^{3,4,24}$. Since no complete topological map of the WWW is available, we limited our study to a subset of the web containing 325,729 nodes and 1,469,680 links ($\langle k \rangle = 4.59$) (ref. 3). Despite the directedness of the links, the response of the system is similar to the undirected networks we investigated earlier: after a slight initial increase, d remains constant in the case of random failures and increases for attacks (Fig. 2c). The network survives as a large cluster under high rates of failure, but the behaviour of $\langle s \rangle$ indicates that under attack the system abruptly falls apart at $f_c^f = 0.067$ (Fig. 3c).

In summary, we find that scale-free networks display a surprisingly high degree of tolerance against random failures, a property not shared by their exponential counterparts. This robustness is probably the basis of the error tolerance of many complex systems, ranging from cells⁸ to distributed communication systems. It also explains why, despite frequent router problems²³, we rarely experience global network outages or, despite the temporary unavailability of many web pages, our ability to surf and locate information on the web is unaffected. However, the error tolerance comes at the expense of attack survivability: the diameter of these networks increases rapidly and they break into many isolated fragments when the most connected nodes are targeted. Such decreased attack survivability is useful for drug design⁸, but it is less encouraging for communication systems, such as the Internet or the WWW. Although it is generally thought that attacks on networks with distributed resource management are less successful, our results indicate otherwise. The topological weaknesses of the current communication networks, rooted in their inhomogeneous connectivity distribution, seriously reduce their attack survivability. This could be exploited by those seeking to damage these systems. □

Received 14 February; accepted 7 June 2000.

- Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, 47–52 (1999).
- Claffy, K., Monk, T. E. *et al.* Internet tomography. *Nature Web Matters* [online] (7 Jan. 99) (<http://helix.nature.com/webmatters/tomog/tomog.html>) (1999).
- Albert, R., Jeong, H. & Barabási, A.-L. Diameter of the World-Wide Web. *Nature* **401**, 130–131 (1999).
- Kumar, R., Raghavan, P., Rajalopagan, S. & Tomkins, A. in *Proc. 9th ACM Symp. on Principles of Database Systems* 1–10 (Association for Computing Machinery, New York, 2000).
- Huberman, B. A. & Adamic, L. A. Growth dynamics of the World-Wide Web. *Nature* **401**, 131 (1999).
- Faloutsos, M., Faloutsos, P. & Faloutsos, C. On power-law relationships of the internet topology, ACM SIGCOMM '99. *Comput. Commun. Rev.* **29**, 251–263 (1999).
- Wasserman, S. & Faust, K. *Social Network Analysis* (Cambridge Univ. Press, Cambridge, 1994).
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* (in the press).
- Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17–60 (1960).
- Bollobás, B. *Random Graphs* (Academic, London, 1985).
- Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
- Zegura, E. W., Calvert, K. L. & Donahoo, M. J. A quantitative comparison of graph-based models for internet topology. *IEEE/ACM Trans. Network* **5**, 770–787 (1997).
- Williams, R. J. & Martinez, N. D. Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000).
- Maritan, A., Colaiori, F., Flammini, A., Cieplak, M. & Banavar, J. Universality classes of optimal channel networks. *Science* **272**, 984–986 (1996).
- Banavar, J. R., Maritan, A. & Rinaldo, A. Size and form in efficient transportation networks. *Nature* **399**, 130–132 (1999).
- Barthélemy, M. & Amaral, L. A. N. Small-world networks: evidence for a crossover picture. *Phys. Rev. Lett.* **82**, 3180–3183 (1999).
- Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–511 (1999).
- Barabási, A.-L., Albert, R. & Jeong, H. Mean-field theory for scale-free random networks. *Physica A* **272**, 173–187 (1999).
- Redner, S. How popular is your paper? An empirical study of the citation distribution. *Euro. Phys. J. B* **4**, 131–134 (1998).
- Lawrence, S. & Giles, C. L. Accessibility of information on the web. *Nature* **400**, 107–109 (1999).
- Milgram, S. The small-world problem. *Psychol. Today* **2**, 60–67 (1967).
- Bunde, A. & Havlin, S. (eds) *Fractals and Disordered Systems* (Springer, New York, 1996).
- Paxson, V. End-to-end routing behavior in the internet. *IEEE/ACM Trans. Network* **5**, 601–618 (1997).
- Adamic, L. A. The small world web. *Lect. Notes Comput. Sci.* **1696**, 443–452 (1999).

Acknowledgements

We thank B. Bunker, K. Newman, Z. N. Oltvai and P. Schiffer for discussions. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to A.-L.B. (e-mail: alb@nd.edu).

Controlled surface charging as a depth-profiling probe for mesoscopic layers

Ilanit Doron-Mor*, Anat Hatzor*, Alexander Vaskevich*, Tamar van der Boom-Moav†, Abraham Shanzer‡, Israel Rubinstein* & Hagai Cohen‡

* Department of Materials and Interfaces, †Department of Organic Chemistry, ‡Department of Chemical Services, Weizmann Institute of Science, Rehovot 76100, Israel

Probing the structure of material layers just a few nanometres thick requires analytical techniques with high depth sensitivity. X-ray photoelectron spectroscopy¹ (XPS) provides one such method, but obtaining vertically resolved structural information from the raw data is not straightforward. There are several XPS depth-profiling methods, including ion etching², angle-resolved XPS (ref. 2) and Tougaard's approach³, but all suffer various limitations^{2–5}. Here we report a simple, non-destructive XPS depth-profiling method that yields accurate depth information with nanometre resolution. We demonstrate the technique using self-assembled multilayers on gold surfaces; the former contain 'marker' monolayers that have been inserted at predetermined depths. A controllable potential gradient is established vertically through the sample by charging the surface of the dielectric overlayer with an electron flood gun. The local potential is probed by measuring XPS line shifts, which correlate directly with the vertical position of atoms. We term the method 'controlled surface charging', and expect it to be generally applicable to a large variety of mesoscopic heterostructures.

Charging is commonly considered an experimental obstacle to accurate determination of binding energies in XPS measurements of poorly conducting surfaces⁶. To compensate the extra positive charge (a natural consequence of photoelectron emission) and to stabilize the energy scale on a reasonably correct value, an electron flood gun is routinely used, creating a generally uniform potential across the studied volume. This, however, is often impossible with systems comprising components that differ in electrical conductivity^{7–9}. In such cases, chemical information may be smeared due to XPS line shifts that follow local potential variations. On the other hand, this very effect can be used to gain structural information^{10–12}. We have shown recently that surface charging can be used to analyse self-assembled monolayers on heterogeneous substrates, providing lateral resolution on a scale given by the substrate structural variations—that is, much smaller than the probe size¹³. Here we show that controlled surface charging (CSC) can be applied to obtain nanometre-scale depth resolution in laterally homogeneous structures.

Well-defined correlation between the local potential and the vertical (depth) axis can be achieved in some systems; for example, those consisting of an insulating overlayer on a metallic (grounded) substrate, where the overlayer is laterally uniform. We use such a system in the work reported here. In the present application the

flood electron gun is not used for charge neutralization, but rather is operated at a considerably higher flux of electrons; this creates a dynamic balance between charge generation and discharge rates, which is controlled by variation of the flood-gun parameters (that is, electron flux and kinetic energy). Evidently, the extra negative charge is accumulated on the film surface, and a dynamic steady state is reached where the charge leakage to the grounded substrate is balanced by the net incoming flux. With the systems studied here, essentially no space charge is developed within the dielectric film. The resultant vertical field is therefore constant, such that the local potential, correlated directly with XPS line shifts, varies linearly with the depth scale (z), unfolding the spectrum with respect to the vertical scale. The steady-state situation, with a surface charge density σ and an overlayer dielectric constant ϵ , is thus modelled by a parallel-plate capacitor, where the local potential ϕ between the plates is given by equation (1) (σ is determined by the overlayer resistivity and the flood-gun operating conditions).

$$\phi(z) = 4\pi\sigma z/\epsilon \quad (1)$$

The studied substrates comprised a 100-nm {111} textured gold layer, evaporated on a highly polished doped Si(111) wafer¹⁴. Metal–organic coordinated multilayers were constructed layer-by-layer on the gold surface as previously described¹⁴. The base layer was a disulphide-bishydroxamate molecule (compound **1** in Fig. 1d)¹⁵. Two different bifunctional ligand repeat units were used, a tetrahydroxamate (**2**, Fig. 1d)¹⁴ and a diphosphonate (**3**, Fig. 1d)^{16–18}, connected by three different tetravalent metal ions: Zr⁴⁺, Ce⁴⁺ or Hf⁴⁺.

XPS measurements were performed on a Kratos Analytical AXIS-HS Instrument, using a monochromatized Al ($K\alpha$) source ($h\nu = 1,486.6$ eV) at a relatively low power¹³. The flood gun was used under varying conditions, of which four are presented here (see Fig. 2 legend). Accurate determination of line shifts was achieved by graphically shifting lines until a statistically optimal match was obtained. This procedure, correlating the full line shape (100–200 data points), allowed excellent accuracy (~ 0.03 eV) in the determination of energy shifts, much beyond the experimental energy resolution (that is, source and bare line widths). In cases where the external charging induced line-shape changes, larger uncertainties were considered. The shifts were translated to potential differences relative to the gold by subtracting the corresponding gold line shifts. Special effort was made to determine time periods characteristic of the stabilization of local potentials, which were generally longer for the lower flood-gun flux. This was of particular importance in view of beam-induced damage⁵, which was observed after an hour or more and found to slightly distort the CSC behaviour. The latter will be discussed elsewhere.

For a general demonstration of CSC, two sets of multilayers with varying thicknesses (2 to 10 layers) were constructed (Fig. 1a), using the tetrahydroxamate ligand **2**. Each set was assembled with a different binding ion, that is, Zr(IV) or Ce(IV), respectively. The spectral response to the flood-gun flux is shown in Fig. 2a. The various surface elements manifest energy shifts and line-shape changes, both associated with the development of vertical potential gradients: the Au 4f line shifts by a very small amount, attributed to the finite (sample-dependent) conductivity of the silicon substrate. All other overlayer elements exhibit much larger shifts, with differences correlated with their spatial (depth) distribution. The energy shifts and line shapes of elements not shown in Fig. 2 (N, O) are fully consistent with the model.

In these two sets of samples, all the overlayer elements (except sulphur) are distributed along the vertical scale. This complicates the derivation of local potentials in general, and specifically the overall potential difference, V_0 , developed across the entire overlayer. (V_0 here is relative to the 'gun-off' situation.) Yet an approximate analysis can be performed by curve-fitting the metal ion lines to sets of discrete signals, corresponding to ion interlayers at

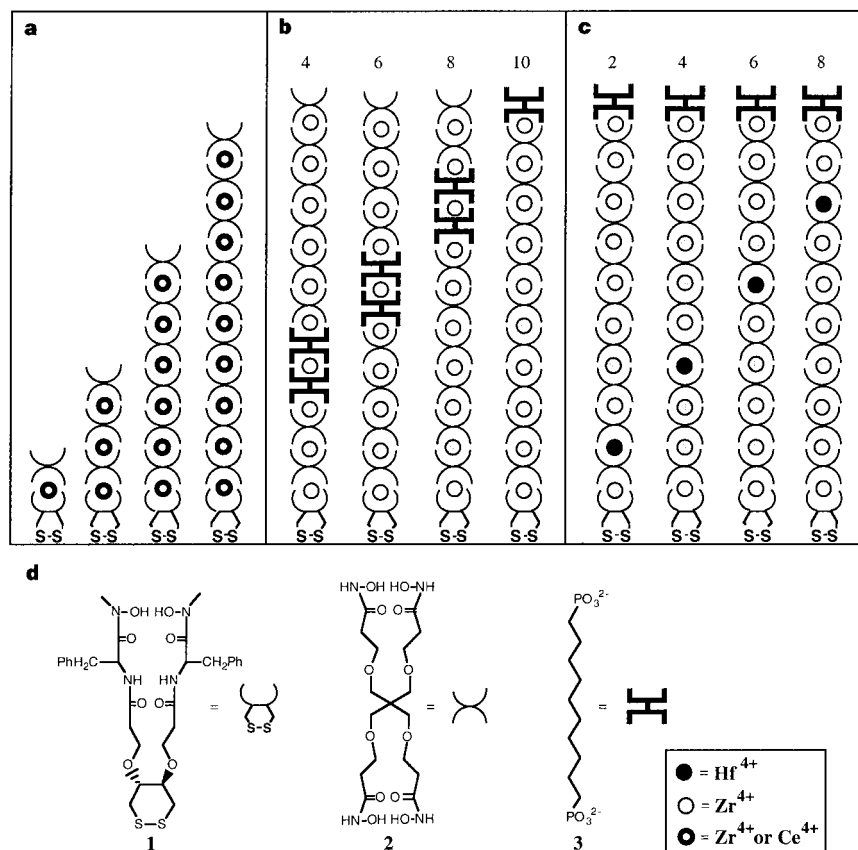


Figure 1 Schematic representation of the studied multilayer systems. In panels **a–c** each ‘molecular column’ represents a full multilayer, bound to the Au substrate through disulphide groups. The bold symbols represent the marker units. In **d** are shown the

meanings of the symbols used in these column diagrams. The numbers above the columns in **b** and **c** indicate the position of the ‘moving’ marker (abscissa in Fig. 4).

progressive vertical positions. The Zr 3d doublet in Fig. 2a indeed exhibits asymmetric broadening under the ‘gun-on’ conditions, in contrast with the ‘gun-off’ situation where potential gradients (of opposite sign) are much smaller, thus having minimal effect on the line shape. (A detailed discussion of potential gradients developed under ‘gun-off’ conditions will not be given here.) The curve fitting of the ‘gun-on’ line (Fig. 2a) yields energy shifts of 0.18 eV per layer and an intensity attenuation factor of 1.72 per layer, in good agreement with further results given below. Hence, even without elaborate curve fitting, the Zr (and similarly, the Ce) line can provide an approximate value of V_0 from the shift of the low-binding-energy side of the line (rather than the peak), corresponding to the uppermost layer with the largest shift. (The potential gradient in the ‘gun-off’ situation is much smaller, ~ 0.035 eV per layer (from curve fitting of the ‘gun-off’ lines, not shown). Consideration of these small gradients would increase the derived V_0 values by $\leq 10\%$.)

The results, presented as V_0 versus multilayer thickness (the former derived from the Zr 3d low-binding-energy side), are shown in Fig. 3. Two important conclusions are drawn from these results. First, the potential difference across the full overlayer, V_0 , increases linearly with the layer thickness, suggesting that the layer is practically free of space charge. The system may, therefore, be satisfactorily simulated by the simple electrical circuit shown in Fig. 3 inset. Second, the absolute gradient values, associated with the vertical film resistivity, are medium-sensitive; the nearly three-fold difference in the slope of V_0 versus film thickness for Zr(IV) and Ce(IV) suggests a marked influence of the binding ion on the electrical conductivity of the layer, an issue of substantial interest but beyond the scope of this work (complementary

characterization, not described here, precludes structural imperfections as a possible cause of conductivity differences). This emphasizes the power of the CSC method in analysing electrical properties of ultrathin films.

We show the enhanced accuracy of CSC depth profiling with two additional sets of tetrahydroxamate-based coordination multilayers, where the energy shift of an indicator atom, installed at a well defined depth, is normalized to V_0 , eliminating the effect of conductivity on the geometric information. Both sets comprised 11-layer films (including the base layer), with Zr(IV) as the binding ion. A full quantitative treatment is achieved when two markers are used (Fig. 1c); the value of V_0 is accurately determined with a diphosphonate layer positioned at the film top, while Hf(IV) serves as the ‘moving’ marker, replacing a Zr(IV) layer at varying depths (a third marker is always the Au substrate). The second set was designed to confirm the compatibility (structural and electrical) of diphosphonate linkers with the tetrahydroxamate multilayer; here diphosphonates, inserted at progressive depths (Fig. 1b)¹⁷, serve as the marker layer, while V_0 is derived (as above, see Fig. 2a) from the shift of the low-binding-energy side of the Zr(IV) lines. Diphosphonate bilayers, rather than monolayers, were used (except for the top position) (Fig. 1b), in order to increase the P signal intensity.

Selected spectra are presented in Fig. 2b. The complete results, summarized in Fig. 4, show linear behaviour of the normalized shifts, yielding values within 5% of the expected geometric positions (10% for the samples with Hf(IV) in position 2). Hence, the energy shifts can be directly converted to depth information on a linear scale and with an accuracy on the order of a single layer thickness, in the present case ~ 1 nm. The results again justify the use of the

simple capacitor model suggested above for the local potential in these systems. Moreover, the agreement between the two sets shows the reliability of the CSC method.

To compare these results with the traditional intensity-based depth profiling, the P, Hf and Au line intensities were quantified. As shown in Fig. 4, good exponential dependence is obtained for the intensity ratios, indicating the high regularity of the studied structures, as previously reported¹⁴. The excellent agreement between the two independent depth-profiling approaches serves to substantiate the CSC results. We note the *in situ* derivation of the attenuation factor obtained from Fig. 4, $d/\lambda = 0.54$, where d is the

single layer thickness and λ the relevant inelastic mean free path (at kinetic energies of 1,300–1,400 eV), calculated here to be 2.73 nm. This value of λ is relatively small compared with values reported for organic media⁶, a result believed to originate from the presence of the metal ions.

Compared to common XPS depth-profiling methods, the CSC-based depth analysis offers several advantages. Ion etching² is inherently destructive and limited in application, particularly with soft matter. Angle-resolved XPS² (ARXPS), considered non-destructive, is hampered when applied to non-planar morphologies, it requires a large number of measurements (which may induce

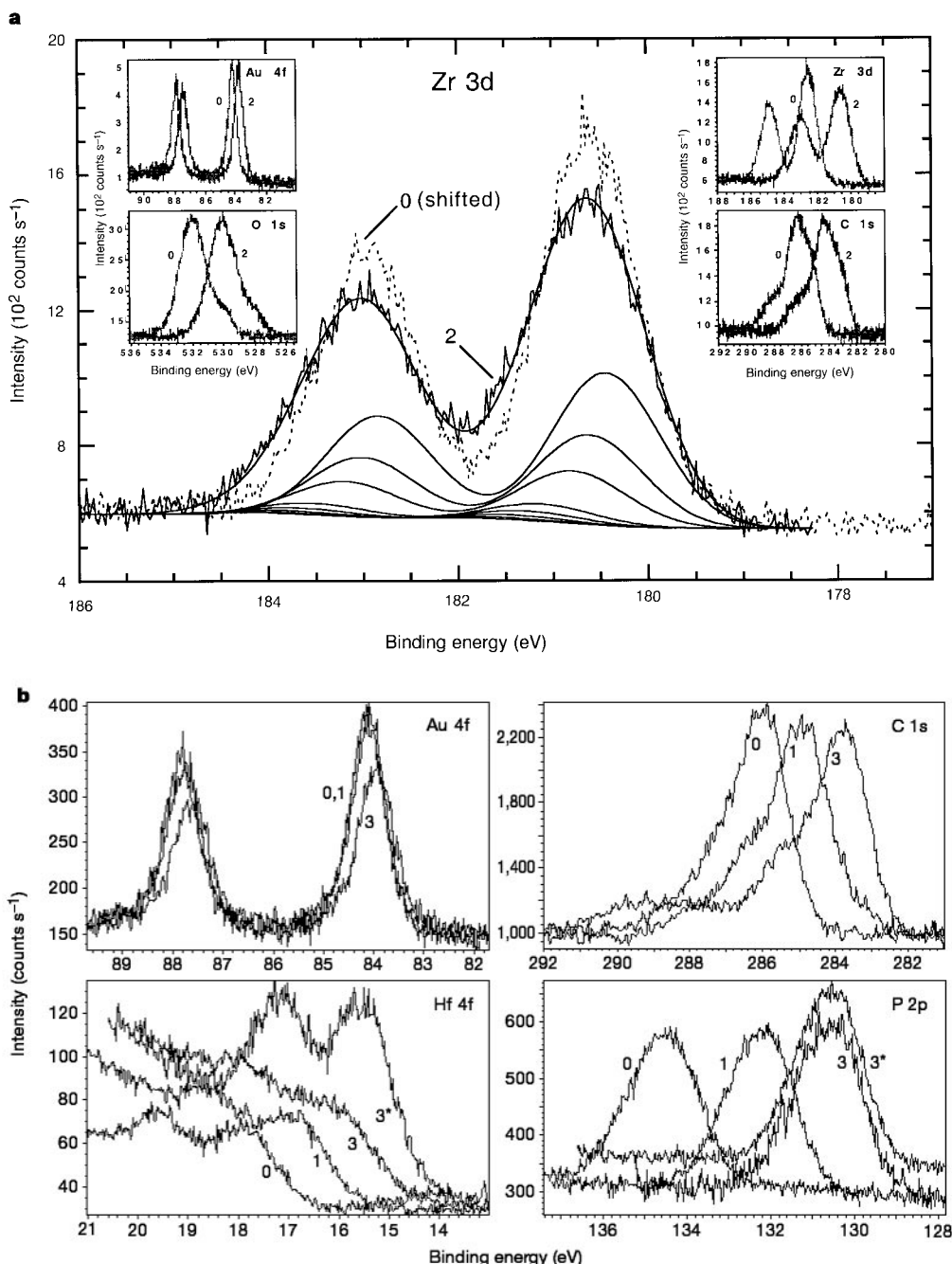


Figure 2 Spectral response to the electron flood-gun flux. **a**, A 10-layer multilayer on gold with Zr(IV) ions (Fig. 1a, right). Line-shape changes of the Zr 3d doublet are shown, comparing flood-gun conditions 0 and 2 (see below), the former shifted by -1.96 eV (dotted line). Curve fitting of 2 ('gun-on') is shown, using three free parameters for the single gaussian and only two parameters (interlayer shift and attenuation) to correlate the nine doublets, corresponding to the discrete ion layers. Insets, raw data of representative

elements at flood-gun conditions 0 and 2. **b**, An 11-layer Zr(IV)-based multilayer with two markers, Hf(IV) and the diphosphonate **3** (Fig. 1c), with the Hf(IV) in layer no. 4. Raw data of representative elements are shown, at flood-gun conditions 0, 1 and 3. Curves 3* correspond to flood-gun condition 3 for a similar multilayer with the Hf(IV) in layer no. 6. Flood-gun conditions are as follows: 0, gun off; 1, 1.90 mA, -1.5 V (emission current and bias voltage, respectively); 2, 1.90 mA, -2.7 V; 3, 1.90 mA, -3.2 V.

damage⁵), and is strongly model-dependent⁴. Tougaard's approach³ (quantitative analysis of signal-to-background correlation) requires minimal interference of neighbouring lines across a wide spectral range, and is therefore less effective with small signals. CSC is basically non-destructive, allowing fast and convenient data collection. It enables differentiation of spectrally identical atoms at different locations. It is applicable to thicker structures than ARXPS, as the signal is not subject to increased attenuation associated with off-normal measurements. Substrate roughness, severely affecting ARXPS depth analysis by smearing the angular scale, would only have a minor effect on CSC depth profiling. The linear dependence on depth, found in the present systems, is an attractive feature of CSC. Note, however, that other systems may involve more elaborate conduction processes, possibly causing deviations from linearity. Such deviations may allow the exploration of additional characteristics of the systems, such as charge distribution and conduction mechanisms.

As a high-resolution depth-profiling method, CSC is applicable to a large variety of non-conducting layers less than about 10 nm thick. Moreover, the present results suggest applications of CSC as a contactless electrical probe, capable of direct detection of local potentials in thin overlayers. In the light of recent progress in enhancing XPS lateral resolution, CSC may be an effective tool

for studying three-dimensional nanostructures and molecular architectures. □

Received 14 March; accepted 26 May 2000.

- Briggs, D. & Seah, M. P. (eds) *Practical Surface Analysis* Vol. 1, 2nd edn (Wiley, New York, 1990).
- Hofmann, S. in *Practical Surface Analysis* Vol. 1, 2nd edn (eds Briggs, D. & Seah, M. P.) 143–199 (Wiley, New York, 1990).
- Tougaard, S. Quantitative analysis of the inelastic background in surface electron spectroscopy. *Surf. Interface Anal.* **11**, 453–472 (1988).
- Tyler, B. J., Castner, D. G. & Ratner, B. D. Regularization—A stable and accurate method for generating depth profiles from angle-dependent XPS data. *Surf. Interface Anal.* **14**, 443–450 (1989).
- Frydman, E., Cohen, H., Maoz, R. & Sagiv, J. Monolayer damage in XPS measurements as evaluated by independent methods. *Langmuir* **13**, 5089–5106 (1997).
- Seah, M. P. in *Practical Surface Analysis* Vol. 1, 2nd edn (eds Briggs, D. & Seah, M. P.) 541–554 (Wiley, New York, 1990).
- Tielsen, B. J., Fulghum, J. E. & Surman, D. J. Differential charging in XPS. 2. Sample mounting and x-ray flux effects on heterogeneous samples. *Surf. Interface Anal.* **24**, 459–468 (1996).
- Barr, T. L. Studies in differential charging. *J. Vac. Sci. Technol. A* **7**, 1677–1683 (1989).
- Lewis, R. T. & Kelley, M. A. Binding energy reference in XPS of insulators. *J. Electron Spectrosc. Relat. Phenom.* **20**, 105–115 (1980).
- Miller, J. D., Harris, W. C. & Zajac, G. W. Composite interface analysis using voltage contrast XPS. *Surf. Interface Anal.* **20**, 977–983 (1993).
- Beamson, G. et al. Characterization of PTFE on silicon wafer tribological transfer films by XPS, imaging XPS and AFM. *Surf. Interface Anal.* **24**, 204–210 (1996).
- Barr, T. L. in *Practical Surface Analysis* Vol. 1, 2nd edn (eds Briggs, D. & Seah, M. P.) 370 (Wiley, New York, 1990).
- Shabtai, K., Rubinstein, I., Cohen, S. & Cohen, H. High-resolution lateral differentiation using a macroscopic probe: XPS of organic monolayers on composite Au-SiO₂ surfaces. *J. Am. Chem. Soc.* **122**, 4959–4962 (2000).
- Hatzor, A. et al. Coordination-controlled self-assembled multilayers on gold. *J. Am. Chem. Soc.* **120**, 13469–13477 (1998).
- Moav, T. et al. Coordination-based symmetric and asymmetric bilayers on gold surfaces. *Chem. Eur. J.* **4**, 502–507 (1998).
- Yang, H. C. et al. Growth and characterization of metal(II) alkanebisphosphonate multilayer thin-films on gold surface. *J. Am. Chem. Soc.* **115**, 11855–11862 (1993).
- Hatzor, A. et al. A metal-ion coordinated hybrid multilayer. *Langmuir* **16**, 4420–4423 (2000).
- Umemura, Y., Yamagishi, A. & Tanaka, K.-I. X-ray photoelectron spectroscopic study of alternately layered zirconium and hafnium phosphate thin films on silicon substrates. *Bull. Chem. Soc. Jpn* **70**, 2399–2403 (1997).

Acknowledgements

This work was supported by the Israel Science Foundation, and the Tashtiot Infrastructure Program of the Israel Ministry of Science.

Correspondence and requests for materials should be addressed to I.R. (e-mail: israel.rubinstein@weizmann.ac.il) or H.C. (e-mail: cphagai@wis.weizmann.ac.il).

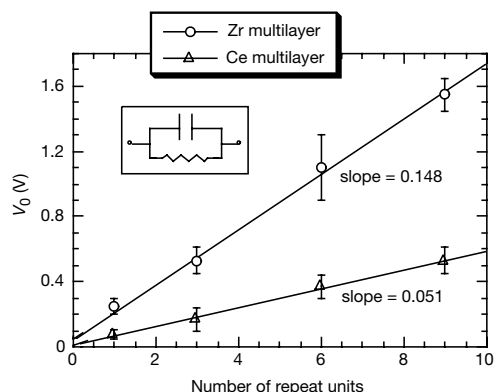


Figure 3 Plot of V_0 versus overlayer thickness. V_0 is the total potential difference across the overlayer, and the overlayer thickness is given as the number of molecular layers. V_0 is derived from the low-binding-energy side of the peaks of the respective ions. Inset, the electrical analogue of the overlayer, with a simple parallel-plate capacitor and a shunt resistor.

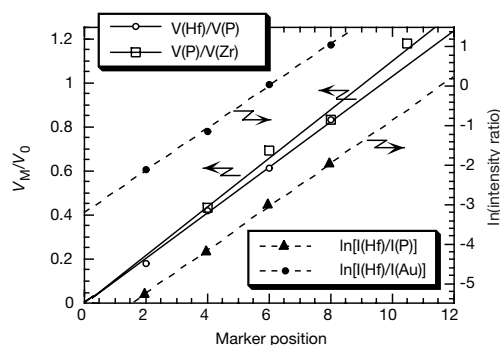


Figure 4 Depth profiling with accurately positioned markers. Solid lines show the normalized local potential of the 'moving' marker, V_M/V_0 , versus marker position, for P and Hf markers in the multilayers shown in Fig. 1b and c, respectively (linear scale). Dashed lines show depth analysis via line intensities for the multilayers shown in Fig. 1c (log scale). The films with Hf in position 2 (Fig. 1c, left) displayed weak Hf signals that required second-derivative data processing for reliable determination of the energy shifts.

Signatures of granular microstructure in dense shear flows

Daniel M. Mueth*, Georges F. Debregeas*, Greg S. Karcmar†, Peter J. Eng‡, Sidney R. Nagel* & Heinrich M. Jaeger*

*The James Franck Institute and Department of Physics, †Radiology Department, ‡Center for Advanced Radiation Sources (CARS), University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA

Granular materials and ordinary fluids react differently to shear stresses. Rather than deforming uniformly, materials such as dry sand or cohesionless powders develop shear bands^{1–5}—narrow zones of large relative particle motion, with essentially rigid adjacent regions. Because shear bands mark areas of flow, material failure and energy dissipation, they are important in many industrial, civil engineering and geophysical processes⁶. They are also relevant to lubricating fluids confined to ultrathin molecular layers⁷. However, detailed three-dimensional information on motion within a shear band, including the degree of particle rotation and interparticle slip, is lacking. Similarly, very little is known about how the microstructure of individual grains affects movement in densely packed material⁵. Here we combine

magnetic resonance imaging, X-ray tomography and high-speed-video particle tracking to obtain the local steady-state particle velocity, rotation and packing density for shear flow in a three-dimensional Couette geometry. We find that key characteristics of the granular microstructure determine the shape of the velocity profile.

In order to probe the role of microstructure inside the narrow granular shear zone, independent determinations are required of the velocity and density profiles with spatial resolution well below the size of individual particles. Non-invasive measurements of this type so far have been limited to two-dimensional (2D) geometries where optical tracking of all particle positions is straightforward^{12,4,8–11}. In a three-dimensional (3D) Couette cell, as sketched in Fig. 1a, a steady-state shear flow can be set up by confining granular material between two concentric, vertical cylinders and turning the inner cylinder at constant velocity v_{wall} while keeping the outer wall at rest. Unlike 2D Couette cells^{9,10,12}, where particles are confined to a single layer with constant volume, there is a free upper surface allowing the packing density to adjust via feedback between shear-induced dilation and gravity.

The difficulty of imaging the interior has restricted studies of 3D granular Couette systems; such studies have probed only the surface (including its velocity fluctuations)¹³, or tracked coloured tracers in very narrow (a few particles wide) cells¹⁴, or measured global quantities, such as the total applied torque. Here we use magnetic resonance imaging (MRI) to obtain flow velocities from the interior of a 3D system^{15,16}. We have used oil-rich seeds as a source of free protons that can be traced using MRI¹⁵. Two kinds of seeds were used to explore the role of microstructure: mustard seeds (spherical with mean diameter $d = 1.8$ mm) and poppy seeds (kidney-shaped

with mean diameter $d = 0.8$ mm). The wall friction was controlled by gluing a layer of seeds on both cylinders. Using a spin-tagging technique, horizontal slices were imaged, as sketched in Fig. 1a. In the resulting MRI image (Fig. 1b), the shear band shows up as the narrow region of deformed stripes near the inner, moving, cylinder. Imaging slices at different heights, h , we measured these deformations, from which the azimuthal velocity profiles were calculated throughout the cell. Similarly, 3D X-ray tomography (Fig. 1d) allowed us to calculate packing-fraction profiles at various heights. At the transparent cell bottom, additional velocity and packing-fraction information was gathered by direct, high-speed-video particle tracking (Fig. 1c).

Before each set of measurements, the cell was run until steady state was reached (as determined by the fact that the packing density profiles became stationary). High-resolution images revealed the long-time average local displacements during a time interval, $\delta t = 100$ ms. Figure 1b represents such displacements averaged over a 17-min acquisition time. The azimuthal velocity at a given distance r from the inner wall was obtained by exploiting the cylindrical symmetry of the Couette geometry: we calculated the MRI intensities along a circle of radius r , $I_{\text{turning}}(\theta)$ and $I_{\text{stopped}}(\theta)$, for the cell turning and at rest, respectively. The position of the central peak in the cross-correlation between $I_{\text{turning}}(\theta)$ and $I_{\text{stopped}}(\theta)$ corresponds to the average azimuthal distance travelled by the material during δt . This technique yielded the angle- and time-averaged radial profile of the steady-state azimuthal mass flow velocity, $v(r)$, resolving $v(r)$ to within 0.1 mm s^{-1} and r to within 0.1 mm. We note that this high-resolution mass flow velocity is the average velocity of all material at radius r and thus not only contains information about particle translation, but also about particle spin. This differs from the

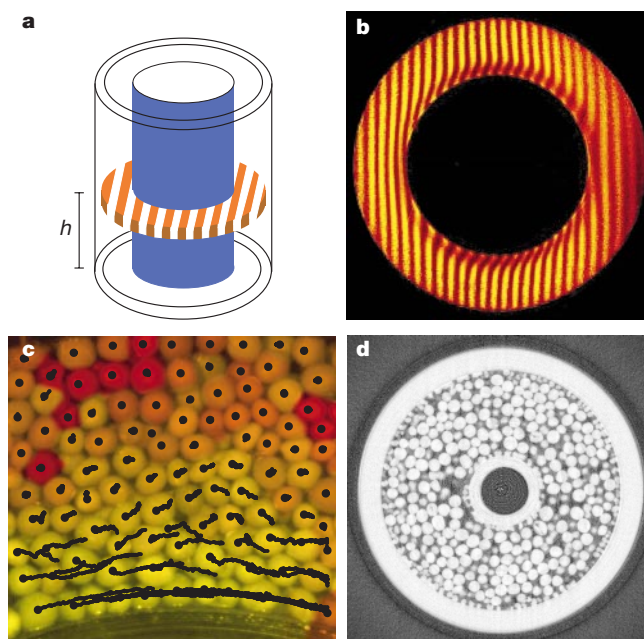


Figure 1 Non-invasive MRI, X-ray tomography and high-speed-video probes of granular Couette flow. **a**, Sketch of the Couette-type shear cell consisting of two concentric cylinders with diameters 51 mm and 82 mm, and filled to a level of 60 mm. The inner cylinder was rotated at velocities $0.6 \text{ mm s}^{-1} < v_{\text{wall}} < 120 \text{ mm s}^{-1}$. The flow velocity was measured using an MRI spin-tagging technique^{14,15}. Before imaging, proton spins were encoded (spin-tagged) so as to display parallel stripes when imaged. Images were taken of 5-mm-thick horizontal slices at various heights, h . **b**, When the inner cylinder was rotated, distortion of the stripe pattern revealed the displacement of the material that occurred during the 100-ms interval between spin-tagging and imaging. 2,048 spin-tag-image steps were used to assemble each complete image. **c**, High-speed-video frame, taken at $1/1,000$ s, of mustard seeds observed through the cell's transparent bottom. To

indicate the movement history of individual particles, their centre positions over the preceding 200 frames are traced by black lines. The particle colouring reflects the magnitude of each particle's average velocity during this 0.2-s interval: fast particles (yellow) near the inner wall appear to move smoothly, while slower particles (orange and red) display more irregular and intermittent motion. In the long-time average, measured by the MRI technique, the flow appears smooth everywhere. **d**, Two-dimensional, horizontal slice through the cell at half the filling height, taken from the X-ray tomography data set. Particles, in this case mustard seeds, appear bright (as do the walls of the cell). These measurements were performed at the Advanced Photon Source at Argonne National Laboratory in a Couette cell of slightly smaller size but under otherwise identical shearing conditions.

average velocity of particle centres that typically is obtained by video particle-tracking techniques.

We found $v(r)$ highly reproducible from run to run, even though on shorter timescales there are rapid velocity fluctuations from point to point within the material that, at the top and bottom boundaries, can be observed with high-speed video (Fig. 1c). Within the resolution of our measurements, $v(r)$ for both types of seeds did not vary with height, including the regions near the top and bottom surfaces. (We note that this indicates that non-azimuthal, secondary flow was small and did not affect $v(r)$, in accordance with tracer-bead studies¹³.) Apart from an overall scale factor, we also did not detect any shear-rate dependence to the velocity profile over the entire range $5 \text{ mm s}^{-1} < v_{\text{wall}} < 120 \text{ mm s}^{-1}$ explored with MRI. (With video imaging we verified rate independence down to velocities as small as 0.6 mm s^{-1}).

For (nearly) monodisperse smooth, spherical particles, the decay of $v(r)$ away from the shearing wall is dominated by abrupt drops at integer multiples of r/d as shown in Fig. 2a for mustard seeds. In the normalized velocity gradient $\gamma(r) \equiv (d/v) \partial v / \partial r$ (Fig. 2b) these drops appear as deep narrow valleys. They are correlated with pronounced oscillations in the packing density, $\rho(r)$, which signal the presence of well defined, single-grain-wide layers near the moving wall (Fig. 2c).

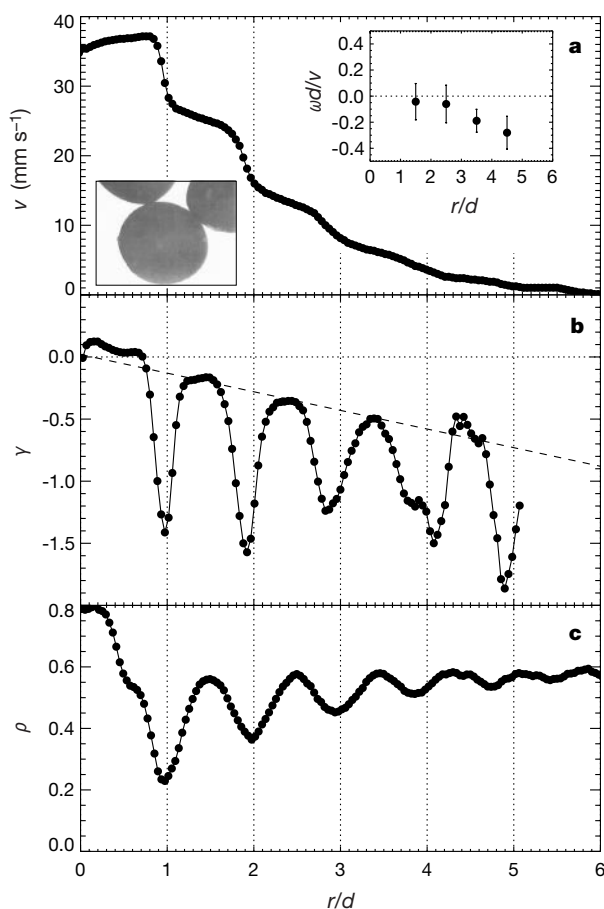


Figure 2 Radial profiles of azimuthal velocity, spin and packing density for spherical mustard seeds. **a**, Steady-state, angle-averaged azimuthal velocity $v(r)$ across the shear band at $h = 30 \text{ mm}$, halfway below the filling level. The layer of seeds glued to the inner cylinder wall extends to the dotted vertical line at $r/d = 1$. The lower inset shows a photograph of the seeds. The upper inset shows the normalized particle spin rate, $\omega d/v$, as a function of distance from the moving wall (where ω is the angular velocity of a particle about its centre). **b**, Normalized velocity gradient $\gamma \equiv (d/v) \partial v / \partial r$ for the data in the main panel of **a**. **c**, Angle-averaged, steady-state radial packing density profile $\rho(r)$ computed from X-ray tomography data as in Fig. 1d, measuring the volume fraction occupied by seed material. For $0 < r/d < 1$, the seeds glued to the wall contribute to ρ .

(We note that the average density approaches the random close packing value at larger r in a manner that is consistent with an exponential form¹⁰.) The shear-induced layering is reminiscent of that seen¹⁷ in the collisional regime of dilute flows, but occurs here in the high-density limit of rate-independent, frictional flow. Because the velocity drops are highly localized, we associate them with slipping at the interface between adjacent layers. Either directly from $v(r)$ or by comparing the integrated areas of each valley in $\gamma(r)$ we find that across each slip zone the velocity decreases by approximately the same factor, $b = 0.36 \pm 0.13$.

In addition to slip between layers, MRI resolves a non-zero velocity gradient $\gamma(r)$ within each layer. This could be caused by either particle rotation or disorder in the layering along the radial direction. (Along the azimuthal and axial directions, particles within layers certainly show packing disorder—as can be seen from the maxima in $\rho(r)$ in Figs 2c and 3c, whose values lie significantly below those for a crystalline configuration). For perfectly arranged layers, $\gamma(r)$ at the layer centres would be determined solely by particle rotation (spin) within each layer. The presence of particles which do not lie perfectly within the layers (revealed by the non-zero values of $\rho(r)$ between layers in Fig. 2c and seen directly from the tracks in Fig. 1c) reduces the gradient in the slip region and increases it at the layer centres. In the limit of complete disorder, we would expect the staircase shape of $v(r)$ to vanish completely. Using information about the density, $\rho_c(r)$, and velocity, $v_c(r)$, of the particle centres, as determined by the high-speed-video and X-ray experiments, we find that approximately 90% of $\gamma(r)$ at the layer centres is due to the radial disorder. In order

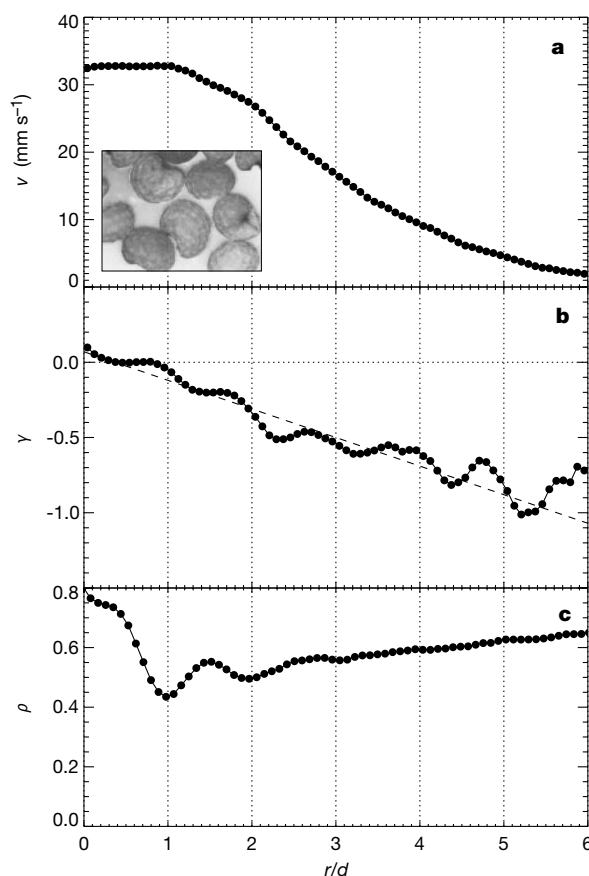


Figure 3 Radial profiles of azimuthal velocity, spin and packing density for aspherical poppy seeds. **a**, Steady-state, angle-averaged azimuthal velocity $v(r)$ across the shear band at $h = 30 \text{ mm}$. The inset shows a photograph of the seeds. **b**, Normalized velocity gradient for the data in **a**. **c**, Corresponding angle-averaged, radial packing density profile $\rho(r)$ from X-ray data. For $0 < r/d < 1$, the seeds glued to the wall contribute to ρ .

for the MRI and high-speed-video experiments to give the same $\gamma(r)$, particle spin must be considered, revealing the spin profile seen in the inset of Fig. 2a. We note that the spin is small, increasing slowly with distance from the shearing wall to a value at $r/d = 4.5$ of less than one full particle rotation per $13d$ translation.

The message of Fig. 2 is that the overall shape of $v(r)$ across the shear band can be understood as arising from two main contributions to the velocity gradient $\gamma(r)$: a slip contribution in the presence of layering, and a second contribution associated with radial disorder. These two pieces are distinguished by the characteristic r -dependence they produce in $v(r)$ on length scales larger than a single grain. A constant slipping fraction, $\int_{\text{interface}} \gamma_{\text{slip}}(r) dr = -b$, across each interface leads (by itself) to a velocity profile with an exponential decay, $v(r) = v_0 \exp(-br/d)$. The constant inter-layer slip appears on a background due to radial disorder, $\gamma_{\text{disorder}}(r) = -2c(r - r_0)/d$, that starts at r_0 within the glued-on layer and increases linearly in strength with slope $2c$ as indicated by the dotted line in Fig. 2b). Such linear r -dependence in $\gamma(r)$ results in a gaussian profile, $v(r) = v_0 \exp(-c(r/d - r_0/d)^2)$. The sum of the two contributions to $\gamma(r)$, after integration, leads to a product of an exponential and gaussian term for the overall, averaged velocity profile: $v(r) = v_0 \exp[-br/d - c(r/d - r_0/d)^2]$. Fits of this function to mustard-seed data yield $b = 0.36 \pm 0.13$, $c = 0.06 \pm 0.03$ and $r_0/d = 0.6 \pm 0.8$ independent of height and shear rate.

The decomposition of $\gamma(r)$ allows the quantitative tracking of the slip inside the shear band for different particle types. We find a greatly reduced slip rate b when smooth spherical particles are replaced by roughened ones. More dramatic differences result from changes in the particle shape. For kidney-shaped poppy seeds we observe a comparatively smooth overall velocity profile (Fig. 3a). Much smaller modulations in $\gamma(r)$ (Fig. 3b) together with little layering in $\rho(r)$ (Fig. 3c) indicate that interlayer slip is greatly reduced. The linear trend $\gamma(r) = -2c(r/d - r_0/d)$ is still present, however, and leads to the gaussian shape for $v(r)$ seen in Fig. 4 over several orders of magnitude. These data also demonstrate explicitly that the shear rate fixed by v_{wall} , while setting the overall scale, leaves the shape of the profile unchanged. This allows for a collapse of all poppy-seed data for different shear rates and heights in the cell onto a single gaussian described by $c = 0.11 \pm 0.02$ and $r_0/d = -0.1 \pm 0.5$. Layering can also be suppressed and radial disorder be promoted in sphere packings if wide particle size distributions are used. We find that these, too, produce essentially gaussian velocity profiles.

From these results, the gaussian component of the velocity profile emerges as a robust, generic feature; slip, if it occurs, is seen to augment $v(r)$ by an additional, exponential factor. In contrast to the translationally invariant slip, the explicit dependence of the gaussian on the distance from the shearing wall indicates correlations. There are several possible mechanisms for radial correlations^{5,9,10,18–20},

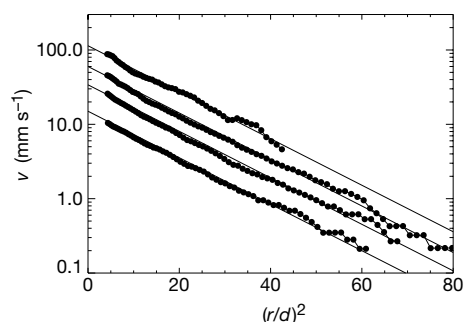


Figure 4 Rate-independence of the velocity profiles. Shown is a log plot of poppy-seed velocity profiles $v(r)$ as a function of $(r/d)^2$ for four different shearing wall speeds ($v_{\text{wall}} = 120, 52, 33$ and 14 mm s^{-1} , top to bottom). Straight lines on this plot correspond to a gaussian with origin at $r = 0$. The slope, c , of these lines corresponds to half the average slope of the velocity gradient in Fig. 3b and characterizes the width $\sigma = d/c^{1/2}$ of the shear band.

including the formation of stress chains^{5,9,10} or particle clusters^{18,20}, both of which require high packing densities and the absence of slip or some degree of interlocking. However, a detailed mechanism that would explain the observed gaussian dependence does not exist at present. A successful theory must not only account for the linear increase in the magnitude of the velocity gradient with r , but also show how the gaussian shape is independent of the presence of layering. We speculate that both of these might be provided by a coupling of $\gamma(r)$ to the average packing fraction and its gradients (for 2D Couette systems, a coupling to the average density alone has been suggested^{9,10}). A gaussian, then, arises naturally for a coupling of type $\gamma(r) \propto \rho(r_0) - \rho(r)$, considering first-order terms in r . Because all stress-loading is driven by the seed layer glued to the inner wall, its location, r_0 , serves as a spatial reference point. This argument carries over to the absence of a gaussian contribution to $v(r)$ in less-dense types of flow, where interlocking may be less effective, such as rapid flows down inclines with free upper surface for which power-law forms have been reported^{14,15,21}, or vertical gravity-driven flows through 2D ‘pipes’, which appear to follow essentially exponential profiles^{2,22}.

Our results show that, in this regime of high packing density and slow shearing rate, there is a direct connection between the microstructure and the shape of the velocity profile. For equal-sized spherical particles, where considerable layering is found, a strong exponential contribution is observed. For aspherical seeds, which show no pronounced layering, the profile is almost completely described by a gaussian centred on the shearing wall. But despite the complicated nature of the grain–grain interactions, the steady-state velocity profile exhibits robust behaviour: its shape—characterized by the two length scales, $\lambda = d/b$ and $\sigma = d/c^{1/2}$ —is found to be independent of both height and shear-rate. □

Received 13 December 1999; accepted 7 June 2000.

1. Bridgwater, J. On the width of failure zones. *Géotechnique* **30**, 533–536 (1980).
2. Nedderman, R. M. & Laohakul, C. The thickness of the shear zone of flowing granular materials. *Powder Technol.* **25**, 91–100 (1980).
3. Mühlhaus, H.-B. & Vardoulakis, I. The thickness of shear bands in granular materials. *Géotechnique* **37**, 271–283 (1987).
4. Drake, T. G. Structural features in granular flows. *J. Geophys. Res.* **95**, 8681–8696 (1990).
5. Oda, M. & Kazama, H. Microstructure of shear bands and its relation to the mechanism of dilatancy and failure of dense granular material. *Géotechnique* **48**, 465–481 (1998).
6. Scott, D. R. Seismicity and stress rotation in a granular model of the brittle crust. *Nature* **381**, 592–595 (1996).
7. Grannick, S. Soft matter in a tight spot. *Phys. Today* **52**, 26–31 (1999).
8. Tüzün, U. & Nedderman, R. M. An investigation of the flow boundary during steady-state discharge from a funnel-flow bunker. *Powder Technol.* **31**, 27–43 (1982).
9. Howell, D., Behringer, R. P. & Veje, C. Stress fluctuations in a 2D granular Couette experiment: a continuous transition. *Phys. Rev. Lett.* **82**, 5241–5244 (1999).
10. Veje, C. T., Howell, D. W. & Behringer, R. P. Kinematics of a two-dimensional granular Couette experiment at the transition to shearing. *Phys. Rev. E* **59**, 739–745 (1999).
11. Rajchenbach, J. Granular flows. *Adv. Phys.* **49**, 229–256 (2000).
12. Karion, A. & Hunt, M. L. Wall stresses in granular Couette flows of mono-sized particles and binary mixtures. *Powder Technol.* **109**, 145–163 (2000).
13. Losert, W., Bocquet, L., Lubensky, T. C. & Gollub, J. Partial dynamics in sheared granular matter. Preprint cond-mat/0004401 at (<http://xxx.lanl.gov>) (24 April 2000).
14. Khosropour, R., Zirinsky, J., Pak, H. K. & Behringer, R. P. Convection and size segregation in a Couette flow of granular material. *Phys. Rev. E* **56**, 4467–4473 (1997).
15. Fukushima, E. Nuclear magnetic resonance as a tool to study flow. *Annu. Rev. Fluid Mech.* **31**, 95–123 (1999).
16. Ehrichs, E. E. *et al.* Granular convection observed by magnetic resonance imaging. *Science* **267**, 1632–1634 (1995).
17. Hopkins, M. A., Jenkins, J. T. & Louge, M. Y. On the structure of three-dimensional shear flows. *Mech. Mater.* **16**, 179–188 (1993).
18. Debregeas, G. & Josserand, C. A self-similar model for shear flows in dense granular material. Preprint cond-mat/9901336 (cited 24 Feb. 2000) at (<http://xxx.lanl.gov>) (1999).
19. Tkachenko, A. & Putkaradze, V. Mesoscopic physics of granular flows. Preprint cond-mat/9912187 (cited 10 Dec. 1999) at (<http://xxx.lanl.gov>) (1999).
20. Josserand, C. A 2D asymmetric exclusion model for granular flows. *Europhys. Lett.* **48**, 36–42 (1999).
21. Hanes, D. M. & Inman, D. L. Observations of rapidly flowing granular-fluid materials. *J. Fluid Mech.* **150**, 357–380 (1985).
22. Pouliquen, O. & Gutfrand, R. Stress fluctuations and shear zones in quasi-static granular chute flows. *Phys. Rev. E* **53**, 557–561 (1996).

Acknowledgements

We thank E. Fukushima, J. Jenkins, C. Josserand, D. Levine, M. Medved, V. Putkaradze,

M. Rivers and A. Tkachenko for discussions, and D. Stockwell for the donation of mustard seeds for the experiment. G.F.D. acknowledges a postdoctoral Grainger Fellowship. This work was supported by an NSF research grant and by the MRSEC program of the NSF.

Correspondence and requests for materials should be addressed to H.M.J. (e-mail: h-jaeger@uchicago.edu).

Oscillatory cluster patterns in a homogeneous chemical system with global feedback

Vladimir K. Vanag, Lingfa Yang, Milos Dolnik, Anatol M. Zhabotinsky & Irving R. Epstein

Department of Chemistry and Volen Center for Complex Systems, MS 015, Brandeis University, Waltham, Massachusetts 02454-9110, USA

Oscillatory clusters are sets of domains in which nearly all elements in a given domain oscillate with the same amplitude and phase^{1–4}. They play an important role in understanding coupled neuron systems^{5–8}. In the simplest case, a system consists of two clusters that oscillate in antiphase and can each occupy multiple fixed spatial domains. Examples of cluster behaviour in extended chemical systems are rare, but have been shown to resemble standing waves^{9–13}, except that they lack a characteristic wavelength. Here we report the observation of so-called ‘localized clusters’—periodic antiphase oscillations in one part of the medium, while the remainder appears uniform—in the Belousov–Zhabotinsky reaction–diffusion system with photochemical global feedback. We also observe standing clusters with fixed spatial domains that oscillate periodically in time and occupy the entire medium, and irregular clusters with no periodicity in either space or time, with standing clusters transforming into irregular clusters and then into localized clusters as the strength of the global negative feedback is gradually increased. By incorporating the effects of global feedback into a model of the reaction, we are able to simulate successfully the experimental data.

The Belousov–Zhabotinsky (BZ) reaction is the best studied chemical oscillator; its dynamics resemble those of many important biological and physical nonlinear oscillators^{14–19}. The reaction consists of the oxidation of malonic acid by bromate, catalysed by metal ions or metallo-complexes in acidic aqueous solution.

Here we study pattern formation in the photosensitive Ru(bpy)₃-catalysed BZ reaction^{20–22} in a thin layer of silica gel^{23,24} with photochemical global negative feedback imposed through illumination (see Fig. 1). The average concentration of Ru(bpy)₃³⁺, Z_{av} , taken over the working area of the gel, is employed to control the intensity I of actinic light (from lamp L2), according to $I = I_{max} \sin^2(g(Z_{av} - Z_t))$, where g is the feedback coefficient and Z_t is a target concentration. The target Z_t is varied by changing the reference voltage (RV) and is set close to the steady state value, Z_{ss} , in such a way that the difference $Z_{av} - Z_t$ is always positive. The angle between polarizers P1 and P2 (that is, $g(Z_{av} - Z_t)$) is not allowed to exceed $\pi/2$. The gain of amplifier A5 is used to control the feedback coefficient g . The actinic light produces bromide ion²¹, which inhibits the oxidation of Ru(bpy)₃²⁺.

We investigate how pattern formation depends on g . We choose the initial reagent concentrations in such a way that without any feedback the system generates bulk oscillations. Various travelling wave patterns are found when g is less than g^* , a value that depends on the initial concentrations of reagents. In most cases,

$1 \times 10^4 \text{ M}^{-1} < g^* < 2 \times 10^4 \text{ M}^{-1}$. When the feedback coefficient is slightly above g^* , either travelling wave patterns or cluster patterns are observed, depending on the initial conditions. The width of this bistability range depends on the initial reagent concentrations.

When g exceeds $2 \times 10^4 \text{ M}^{-1}$, standing clusters arise. Figure 2 depicts several patterns of such clusters. Figure 2a shows one period of oscillation of a pattern that arose from initial conditions created by brief illumination of the gel through a striped mask. At $t = 4 \text{ s}$, Z_{av} reaches a maximum. After this instant, the white domains start to fade and the system gradually evolves to the uniform reduced (dark) state. During the second half-period, the regions that were dark during the first half-period become bright, and at $t = 29 \text{ s}$ the pattern displays another maximum in Z_{av} . As adjacent domains oscillate in antiphase, the period, T_{av} , of oscillations of Z_{av} is half the period of the local oscillations. Figures 2b and 2c show patterns that developed from a spiral wave and from uniform initial conditions, respectively. Other patterns of standing clusters can be obtained by varying the initial conditions. Standing clusters do not possess intrinsic spatial periodicity, because the global feedback is sensitive only to the spatially integrated characteristics of the patterns.

At larger g , two other types of patterns arise. If the initial reagent concentrations lie far from the boundary of the oscillatory region of the parameter space, we observe a transition from standing clusters

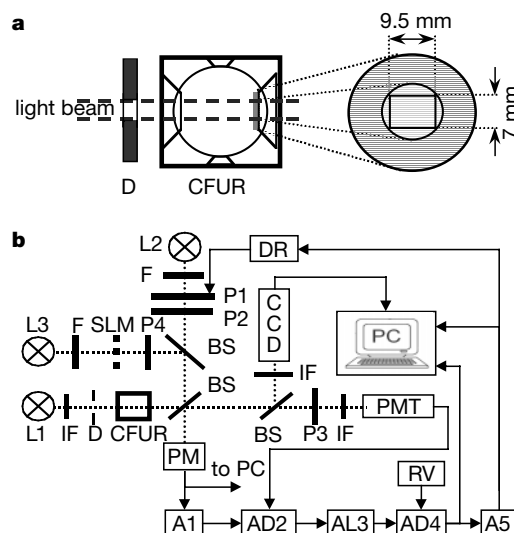


Figure 1 Experimental set-up. **a**, The continuously fed unstirred reactor (CFUR) consists of a 24-ml continuously stirred tank reactor thermostated at 25 °C and a thin layer of silica gel polymerized on the reactor optical window (grey circle). The 0.413-mm-thick gel has diameter 22 mm, and contains 1.5 mM of immobilized Tris(2,2'-bipyridyl)ruthenium(II)(Ru(bpy)₃)²⁺. Diaphragm D selects an illuminated working area of gel (light grey circle). The rectangular frame shows the field of view of the CCD camera. A peristaltic pump injects stock solutions of KBr, malonic acid, NaBrO₃ and H₂SO₄ into a pre-mixing chamber (not shown) and then into the CFUR. **b**, Low-intensity analysing light from the stabilized 45-W light source (L1) passes through the working area of the gel, and is collected by a lens and directed to the photomultiplier tube (PMT). AD2 and AD4 are differential amplifiers, AL3 is a logarithmic amplifier, A5 is a d.c. amplifier. The driver (DR) rotates polarizer P1, and controls the intensity of actinic light from the 450-W Xe arc lamp (L2). P1–P4 are polarizers, IFs are interference filters with $\lambda_{max} = 450 \text{ nm}$, Fs are bandpass filters (400–500 nm), BSs are beam splitters; lenses and collimators are not shown. The 150-W Xe arc lamp (L3) serves to set patterns of initial conditions. An image of the spatial light modulator (SLM) is focused in the plane of the silica gel. Actinic light intensity (I) is measured by the power meter (PM). Polarizers P2–P3 and P4–P3 are crossed to separate optical channels. The remaining parasitic signal from scattered actinic light is compensated electronically by amplifiers A1 and AD2, and stays below 1% of the photomultiplier working signal when I varies from 0 to $I_{max} = 4.3 \text{ mW cm}^{-2}$. Uniformity of illumination of the working area is better than 2%.

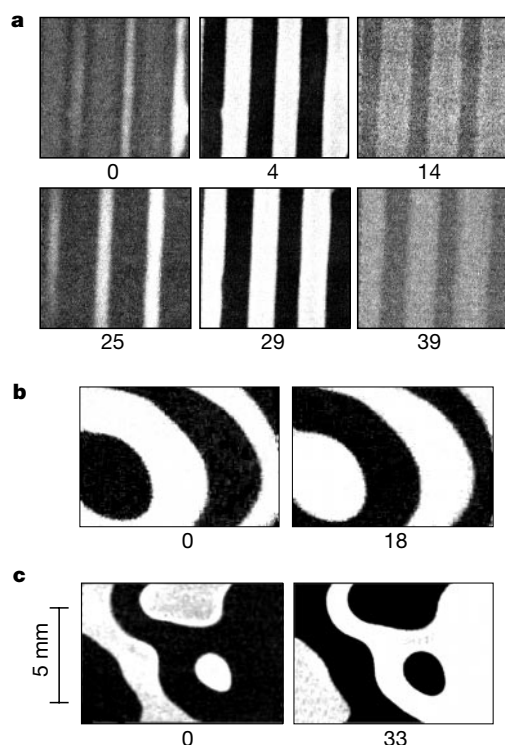


Figure 2 Standing clusters. **a**, One period of oscillation of the pattern formed by initial constant illumination through a striped mask. **b**, The pattern that evolves from an initial spiral wave. Snapshots are taken at two consecutive moments of maximum Z_{av} . **c**, Standing clusters obtained from uniform initial conditions—the gel area was illuminated with constant light to suppress pattern formation, then feedback was switched on and simultaneously constant illumination was switched off. Grey levels quantify $[Ru(bpy)_3^{3+}]$, with white corresponding to maximum and black to minimum. Numbers under frames show time in seconds starting from the beginning (maximum or minimum Z_{av}) of an arbitrary cycle in the stationary regime. Initial reagent concentrations (M) and feedback coefficient (M^{-1}) are: **a**, $[H_2SO_4]_0 = 0.7$, $[BrO_3^-]_0 = 0.28$, $[MA]_0 = 0.2$, $[Br^-]_0 = 0.125$, $g = 2.7 \times 10^4$; **b**, $[H_2SO_4]_0 = 0.7$, $[BrO_3^-]_0 = 0.4$, $[MA]_0 = 0.2$, $[Br^-]_0 = 0.125$, $g = 5.5 \times 10^4$; **c**, $[H_2SO_4]_0 = 0.5$, $[BrO_3^-]_0 = 0.25$, $[MA]_0 = 0.3$, $[Br^-]_0 = 0.1$, $g = 5.5 \times 10^4$. Residence time of fluid in the CSTR is 1,000 s.

to a phenomenon we dub irregular clusters. Local oscillations in these patterns are aperiodic, but the average concentration, Z_{av} , (and, consequently, I) oscillates approximately periodically, and the period, T_{av} , provides a convenient scale for analysing the dynamics of irregular clusters (see Fig. 3b). Figure 3a displays snapshots of irregular clusters separated by T_{av} . Using image analysis techniques we compare the locations of the white domains at successive maxima of Z_{av} . We find that the white domains of two consecutive patterns do not overlap, which indicates antiphase oscillations. The overlapping area of white domains in frames separated by $2T_{av}$ is relatively small, which demonstrates the absence of temporal periodicity. Over a full standing cluster cycle ($2T_{av}$), the white domains of antiphase oscillations cover the entire area of the system. In an irregular cluster, only part of the medium is occupied by the white domains of two consecutive patterns over $2T_{av}$. Because the irregular cluster patterns are not stationary, the entire area is eventually covered by the white domains, typically over several tens of periods.

Localized clusters arise if the initial reagent concentrations are close to the parameter-space boundary between the oscillatory and the reduced steady-state regions. Domains of antiphase oscillations in localized clusters occupy only part of the area, while no pattern can be seen in the remaining part of the system. In Fig. 3c, two snapshots separated by half a period of oscillations show two adjacent large domains of antiphase oscillations separated by a

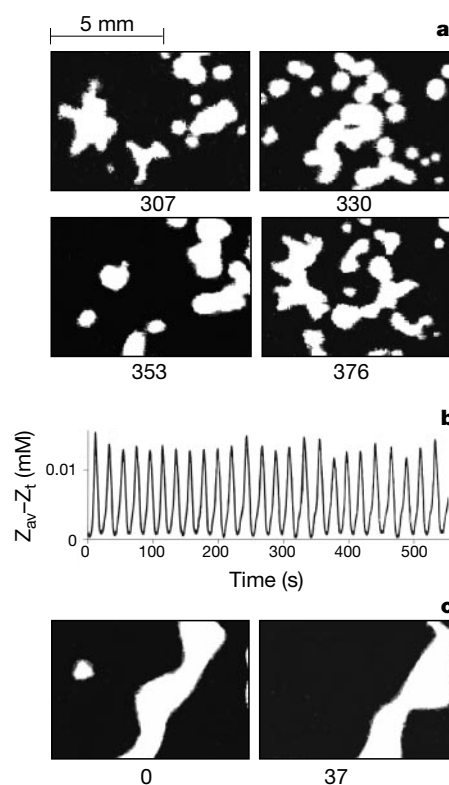


Figure 3 Irregular and localized clusters. Snapshots are taken at consecutive maxima of average concentration of $Ru(bpy)_3^{3+}$ (Z_{av}). **a**, Irregular clusters. Initial concentrations (M): $[H_2SO_4]_0 = 0.75$, $[BrO_3^-]_0 = 0.312$, $[MA]_0 = 0.375$, $[Br^-]_0 = 0.125$, residence time is 667 s; $g = 1.5 \times 10^5 M^{-1}$. **b**, Oscillations of Z_{av} corresponding to **a**. **c**, Localized clusters. Initial concentrations (M): $[H_2SO_4]_0 = 0.5$, $[BrO_3^-]_0 = 0.25$, $[MA]_0 = 0.3$, $[Br^-]_0 = 0.1$ (same as in Fig. 2c); residence time is 1,000 s; $g = 2.2 \times 10^5 M^{-1}$. Other conditions as in Fig. 2.

nodal line. In the first snapshot ($t = 0$), there are also two small domains at the right boundary of the frame and one small domain without the left boundary. The latter domain is within the area with no visible pattern.

With increasing feedback coefficient, the portion of the medium occupied by clusters shrinks. Eventually, clusters disappear, giving way to small-amplitude bulk oscillations. This transition takes place in the range $2 \times 10^5 M^{-1} < g < 3.5 \times 10^5 M^{-1}$, depending on the initial reagent concentrations.

To simulate the pattern formation observed in our experiments, we employ a model of the BZ reaction^{25,26} with the rate constants adjusted according to ref. 27. We add a global linear feedback term to account for the bromide ion production that results from the actinic illumination, $v_{GF} = g\varphi I_{max}(Z_{av} - Z_{ss})$, where φ is the quantum yield. The results of our simulations mimic those of the experiments. Bulk oscillations and travelling waves are observed in the model for smaller values of g . Spontaneous formation of spiral waves predominates at low g . Standing clusters emerge for larger values of the feedback coefficient. Depending on the distance from the boundary between the oscillatory and reduced steady-state parameter regions, we obtain sequences of standing, irregular and localized clusters as shown in Fig. 4 (parameters close to the Hopf bifurcation line) or standing and irregular clusters only (parameters further from the Hopf line). In the simulations, standing clusters arise at lower values of g , irregular clusters at intermediate g , localized clusters at larger g , and small-amplitude bulk oscillations at the largest g , as in the experiments. We also performed simulations with the nonlinear feedback term employed in the experiments and obtained qualitatively the same results. The only

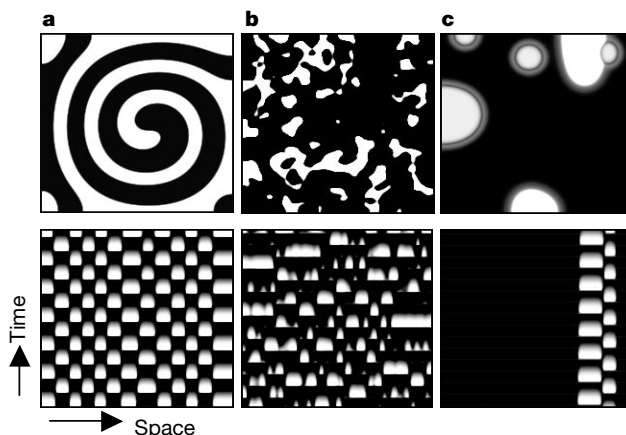


Figure 4 Oscillatory cluster patterns in our BZ reaction–diffusion model with global negative feedback. **a**, Standing clusters: $g = 2 \times 10^4 \text{ M}^{-1}$. **b**, Irregular clusters: $g = 5 \times 10^4 \text{ M}^{-1}$. **c**, Localized clusters: $g = 1.1 \times 10^5 \text{ M}^{-1}$. Top frames display snapshots of patterns; bottom frames show spatio-temporal behaviour along bottom left–top right diagonal of corresponding squares during 380 s. System size is $1 \times 1 \text{ cm}^2$ (200×200 grid points). The initial condition for standing clusters is a spiral wave. Irregular and localized clusters are obtained from the pattern shown in **a** by instantaneous increase of feedback coefficient. Simulation is based on a two-variable model: $\partial X/\partial t = (-k_2X + k_3A)/(k_2X + k_3A)[qk_7k_8BZ/(k_8 + k_{-7}h_0(C-Z)) + k_9B + v_{GF}] - 2k_4X^2 + k_5h_0AX - k_{-5}U^2 + D_X\partial^2X/\partial r^2$, $\partial Z/\partial t = k_6U(C-Z) - k_{-6}XZ - k_7k_8BZ/[k_8 + k_{-7}h_0(C-Z)] + D_Z\partial^2Z/\partial r^2$, where

$U = -k_6(C-Z)/(4k_{-5}) + [k_6(C-Z)]^2 + 8k_{-5}X(2k_5h_0A + k_{-6}Z)]^{1/2}/(4k_{-5})$. Here X , Z , U , A and B are concentrations of HBrO_2 , $\text{Ru}(\text{bpy})_3^{3+}$, HBrO_2^+ , HBrO_3 and malonic acid, respectively. C is the total concentration of catalyst, $[\text{Ru}(\text{bpy})_3^{3+}] + [\text{Ru}(\text{bpy})_3^{2+}]$, h_0 is the Hammett acidity function, and D_X and D_Z are diffusion coefficients. v_{GF} is the global feedback term. Rate constants and parameters are: $k_2 = 2.0 \times 10^6 \text{ M}^{-2} \text{ s}^{-1}$, $k_3 = 2 \text{ M}^{-2} \text{ s}^{-1}$, $k_4 = 3 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$, $k_5 = 33 \text{ M}^{-2} \text{ s}^{-1}$, $k_{-5} = 4.2 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$, $k_6 = 4.0 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$, $k_{-6} = 3.0 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}$, $k_7 = 9.2 \times 10^{-1} \text{ M}^{-1} \text{ s}^{-1}$, $k_8/k_{-7} = 2.5 \times 10^{-4} \text{ M}^2$, $k_9 = 3.0 \times 10^{-6} \text{ s}^{-1}$, $A = 0.278 \text{ M}$, $B = 0.45 \text{ M}$, $C = 2.0 \times 10^{-3} \text{ M}$, $h_0 = 0.25 \text{ M}$, $q = 0.7$, $D_X = 1.5 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$, $D_Z = 2.0 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$ and $\phi_{\text{max}} = 1 \times 10^{-5} \text{ M s}^{-1}$.

exception was that we found only transient localized clusters, which maintained their positions for just a few periods of oscillation.

Time–space plots of the one-dimensional sections of the system in Fig. 4 clarify the dynamics of the system. Irregular clusters are aperiodic in time on a scale at least an order of magnitude larger than the period of oscillation of standing or localized clusters. Our simulations also reveal that the region of travelling wave patterns often overlaps with a part of the standing cluster region, forming a region of bistability on the g -axis, as found in the experiments. The simulations show that diffusion determines a minimum linear size for the domains and a maximum curvature for the nodal lines.

As the global negative feedback gradually increases, the following sequence of patterns is found in our experiments and simulations: bulk oscillations, wave patterns, standing clusters, irregular clusters, localized clusters, small-amplitude bulk oscillations. We suggest that this sequence is fundamental and will be found in other chemical and biological extended systems with global coupling. The discovery of localized clusters is of particular significance, because an enormous variety of patterns of this type can be created from different initial conditions, and their localized character makes them more convenient for information storage and retrieval than patterns that occupy the entire system. Thus they appear to be well suited for natural systems of distributed memory. Cluster formation in models of neural networks with negative global coupling has received considerable attention^{5,7,8}, but the experimental observation of such patterns is difficult. We hope that our findings will stimulate a search for analogous dynamic patterns in natural neural systems. □

Received 11 January; accepted 22 May 2000.

- Golomb, D., Hansel, D., Shraiman, B. & Sompolinsky, H. Clustering in globally coupled phase oscillators. *Phys. Rev. A* **45**, 3516–3530 (1992).
- Hakim, V. & Rappel, W.-J. Dynamics of the globally coupled complex Ginzburg–Landau equation. *Phys. Rev. A* **46**, 7347–7350 (1992).
- Falcke, M., Engel, H. & Neufeld, M. Cluster formation, standing waves, and stripe patterns in oscillatory active media with local and global coupling. *Phys. Rev. E* **52**, 763–771 (1995).
- Wang, W., Kiss, I. Z. & Hudson, J. L. Experiments on arrays of globally coupled chaotic electrochemical oscillators: Synchronization and clustering. *Chaos* **10**, 248–256 (2000).
- Golomb, D. & Rinzel, J. Clustering in globally coupled inhibitory neurons. *Physica D* **72**, 259–282 (1994).

- Terman, D. & Wang, D. Global competition and local cooperation in a network of neural oscillators. *Physica D* **81**, 148–176 (1995).
- Huerta, R., Bazhenov, M. & Rabinovich, M. I. Clusters of synchronization and bistability in lattices of chaotic neurons. *Europhys. Lett.* **43**, 719–724 (1998).
- Rabinovich, M. I., Varona, P. & Torres, J. J. Slow dynamics and regularization phenomena in ensembles of chaotic neurons. *Physica A* **263**, 405–414 (1999).
- Lev, O., Sheintuch, M., Pismen, L. M. & Yarnitzky, Ch. Standing and propagating wave oscillations in the anodic dissolution of nickel. *Nature* **336**, 458–459 (1988).
- Cordonier, G. A., Schuth, F. & Schmidt, L. D. Oscillations in methylamine decomposition on Pt, Rh, and Ir: Experiments and models. *J. Chem. Phys.* **91**, 5374–5386 (1989).
- Jakubith, S., Rotermund, H. H., Engel, W., von Oertzen, A. & Ertl, G. Spatiotemporal concentration patterns in a surface reaction: Propagating and standing waves, rotating spirals, and turbulence. *Phys. Rev. Lett.* **65**, 3013–3016 (1990).
- Somani, M., Liauw, M. A. & Luss, D. Evolution and impact of temperature patterns during hydrogen oxidation on a Ni ring. *Chem. Eng. Sci.* **52**, 2331–2341 (1997).
- Petrov, V., Ouyang, Q. & Swinney, H. L. Resonant pattern formation in a chemical system. *Nature* **388**, 655–657 (1997).
- Belousov, B. P. in *Collection of Short Papers on Radiation Medicine* 145–152 (Medgiz, Moscow, 1959).
- Zhabotinsky, A. M. Periodic liquid phase reactions. *Proc. Acad. Sci. USSR* **157**, 392–395 (1964).
- Zaikin, A. N. & Zhabotinsky, A. M. Concentration wave propagation in a two-dimensional, liquid phase self-oscillating system. *Nature* **225**, 535–537 (1970).
- Winfree, A. T. Spiral waves of chemical activity. *Science* **175**, 634–636 (1972).
- Ross, J., Müller, S. C. & Vidal, C. Chemical waves. *Science* **240**, 460–465 (1988).
- Kapral, R. & Showalter, K. (eds) *Chemical Waves and Patterns* (Kluwer, Dordrecht, 1995).
- Kuhnert, L., Agladze, K. I. & Krinsky, V. I. Image processing using light-sensitive chemical waves. *Nature* **337**, 244–247 (1989).
- Grill, S., Zykov, V. S. & Müller, S. C. Feedback-controlled dynamics of meandering spiral waves. *Phys. Rev. Lett.* **75**, 3368–3371 (1995).
- Kadar, S., Amemiya, T. & Showalter, K. Reaction mechanism for light sensitivity of the $\text{Ru}(\text{bpy})_3^{3+}$ -catalyzed Belousov–Zhabotinsky reaction. *J. Phys. Chem. A* **101**, 8200–8206 (1997).
- Yamaguchi, T., Kuhnert, L., Nagy-Ungvári, Zs., Müller, S. C. & Hess, B. Gel systems for the Belousov–Zhabotinsky reaction. *J. Chem. Phys.* **95**, 5831–5837 (1991).
- Amemiya, T., Kettunen, P., Kadar, S., Yamaguchi, T. & Showalter, K. Formation and evolution of scroll waves in photosensitive excitable media. *Chaos* **8**, 872–878 (1998).
- Zhabotinsky, A. M., Buchholtz, F., Kiyatkin, A. B. & Epstein, I. R. Oscillations and waves in metal-ion-catalyzed bromate oscillating reactions in highly oxidized states. *J. Phys. Chem.* **97**, 7578–7584 (1993).
- Bugrim, A., Dolnik, M., Zhabotinsky, A. M. & Epstein, I. R. Heterogeneous sources of target patterns in reaction–diffusion systems. *J. Phys. Chem.* **100**, 19017–19022 (1996).
- Gao, Y. & Försterling, H.-D. Oscillations in the bromomalonic acid/bromate system catalyzed by $[\text{Ru}(\text{bipy})_3]^{3+}$. *J. Phys. Chem.* **99**, 8638–8644 (1995).

Acknowledgements

This work was supported by the Chemistry Division of the National Science Foundation and the W. M. Keck Foundation.

Correspondence and requests for materials should be addressed to A.M.Z. (e-mail: zhabotinsky@brandeis.edu)

Massive dissociation of gas hydrate during a Jurassic oceanic anoxic event

Stephen P. Hesselbo*, Darren R. Gröcke*, Hugh C. Jenkyns*, Christian J. Bjerrum†, Paul Farrimond‡, Helen S. Morgans Bell* & Owen R. Green*

* Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK

† Danish Centre for Earth System Science, University of Copenhagen, Juliane Maries Vej 30, DK-2100, Denmark

‡ Fossil Fuels and Environmental Geochemistry, Drummond Building, University of Newcastle, Newcastle-Upon-Tyne NE1 7RU, UK

In the Jurassic period, the Early Toarcian oceanic anoxic event (about 183 million years ago) is associated with exceptionally high rates of organic-carbon burial, high palaeotemperatures and significant mass extinction^{1–4}. Heavy carbon-isotope compositions in rocks and fossils of this age have been linked to the global burial of organic carbon, which is isotopically light. In contrast, examples of light carbon-isotope values from marine organic matter of Early Toarcian age have been explained principally in terms of localized upwelling of bottom water enriched in ¹²C versus ¹³C (refs 1,2,5,6). Here, however, we report carbon-isotope analyses of fossil wood which demonstrate that isotopically light carbon dominated all the upper oceanic, biospheric and atmospheric carbon reservoirs, and that this occurred despite the enhanced burial of organic carbon. We propose that—as has been suggested for the Late Palaeocene thermal maximum, some 55 million years ago⁷—the observed patterns were produced by voluminous and extremely rapid release of methane from gas hydrate contained in marine continental-margin sediments.

The better-known positive carbon-isotope excursion of the Early Toarcian is well illustrated by European organic-poor marine carbonates (Figs 1 and 2), whereas the negative excursion is recorded primarily in marine organic-matter and associated carbonate (the latter previously considered to be largely of diagenetic origin²). The negative excursion is also present, albeit very rarely, in Tethyan limestones lacking organic enrichment⁸. Viewed together, these $\delta^{13}\text{C}$ curves depict a period of gradual rise in isotopic values which is interrupted by first a negative excursion, and then continued immediately afterwards as a positive excursion (Fig. 2).

Details of the negative excursion in marine organic matter are provided by a section at Hawsker Bottoms, Yorkshire, England (Fig. 3). There, the relationship of the excursion to black-shale

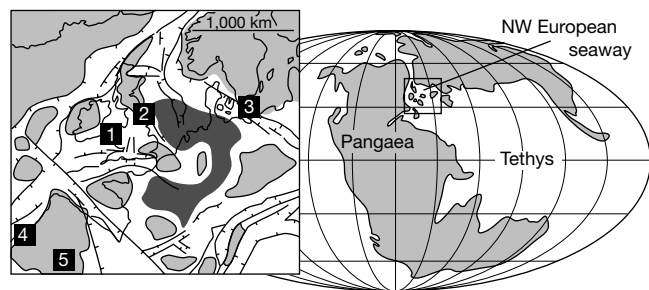


Figure 1 Palaeogeographic maps for Toarcian sections in northwestern Europe^{19,28}. Land areas are shown in medium grey; organic-rich shale^{1,2} in dark grey. 1, Wales, Mochras Farm borehole; 2, England, Hawsker Bottoms; 3, Denmark, Bornholm; 4, Portugal, Porto de Mós; 5, Spain, Fuente de la Vidriera.

deposition and high-resolution ammonite biostratigraphy is simply determined. Fossil wood is also present, preserved as coal (some pyritized) and 'jet', a form of coal that will take a high polish. Carbon-isotope values for wood ($\delta^{13}\text{C}_{\text{wood}}$) and marine organic matter ($\delta^{13}\text{C}_{\text{org}}$) show a close parallelism through the sampled upper *tenuicostatum* and lower *falciferum* zones (Fig. 3). The negative $\delta^{13}\text{C}$ excursion and high total organic carbon (TOC) characterize the *exaratum* subzone of the *falciferum* zone.

Jet has previously been rejected as a monitor of atmospheric carbon-isotope composition because it is known that marine organic carbon has been incorporated during diagenesis⁵. To assess the isotopic effects of this process, six samples were subjected to solvent extraction and pyrolysis-gas chromatography-mass spectrometry (Py-GC-MS) of the residues (a method similar to ref. 9). Extractable organic-matter content ranges from 6.4% to 18.8% and is closely correlated with TOC (Fig. 3). The phenolic and naphthalenic composition of the residue confirms terrestrial (wood) dominance. Extracted bitumen could not account for the isotopic excursion in raw samples because impregnation would have to have been more than 50% to move pure wood values away from a –25‰ baseline to the observed extent. Confirmation comes from carbon-isotope analyses of the residues, which show only minor differences from unextracted samples. We conclude that a primary stratigraphic signal is retained in the fossil wood, despite bitumen contamination from the enclosing shale.

In Yorkshire, the up-section trend towards more negative $\delta^{13}\text{C}_{\text{org}}$ begins at the same level as that where TOC values start to rise, but the isotopic values show an abrupt decrease about 2.4 m above this level (Fig. 3). Time represented in this section has been estimated previously based on an average 50- μm thick lamina-couplet (assumed to be annual)^{10,11}. The overall shift to lower $\delta^{13}\text{C}_{\text{org}}$

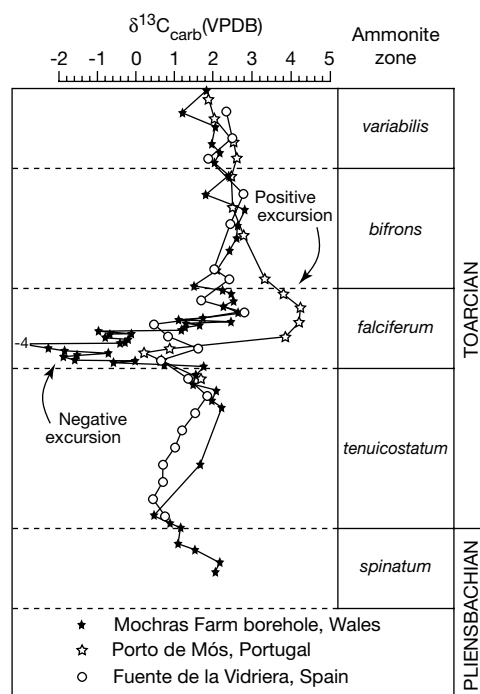


Figure 2 Carbonate carbon-isotope data through selected European Toarcian sections. Wales, Mochras Farm borehole (ref. 2 and new data); Portugal, Porto de Mós²⁹; Spain, Fuente de la Vidriera⁸. See Fig. 1 for locations. We note that $\delta^{13}\text{C}_{\text{carb}}$ values from belemnites can reach up to +6‰ in the upper *falciferum* zone at the location shown in Fig. 3 (compare ref. 2). Tethyan zonal divisions are correlated with those of northern Europe using the first occurrence of the ammonite *Hildaites*¹⁴ as a time-synchronous marker. VPDB, Vienna Pee-Dee belemnite.

values occurs through a thickness of strata representing a minimum duration of about 100 kyr (some laminae may have been lost to non-deposition or erosion) and most of this change takes place between levels b and c in Fig. 3, representing ~70 kyr. Importantly, the 2‰ jump across less than 22 cm of strata at level b (Fig. 3) may represent a much shorter interval, presumably less than about 5 kyr if there is no major hiatus. The decay pattern of the negative anomaly from its peak at level c to the top of the section analysed closely approximates an exponential curve, with relative stability attained after about a further 80 kyr. Thus, close similarities exist with patterns of timing described for the Late Palaeocene thermal maximum (LPTM)¹².

We also observe similar isotopic patterns in Toarcian wood from a near-shore setting, within the Bagå Formation, Bornholm, eastern Danish Basin¹³ (Figs 1 and 4). Samples analysed comprise two sets: (1) macroscopic wood, recognizable to the naked eye, up to branch-sized; and (2) microscopic wood (0.25 to 1 mm average dimension). Where possible, coal (hard, black, unstructured, higher density) and charcoal (soft, black, structured, lower density) were identified and analysed separately. Apart from wood and leaf cuticle there is little organic matter present in the section—the sediments comprise

sand, silt and mud—and thus there is no possibility of contamination by marine carbon.

A striking feature of the $\delta^{13}\text{C}_{\text{wood}}$ data from Bornholm is the major negative excursion (Fig. 4). The $\delta^{13}\text{C}_{\text{wood}}$ values are not influenced by the nature of the preservation (that is, coal versus charcoal). Data from macrofossils and microfossils plot along identical stratigraphic trends, except between 44 and 46 m, where palynological evidence indicates an Aalenian age. Only at this higher level is it likely that the microscopic wood is reworked. Identification of the negative excursion allows precise correlation with the ammonite biostratigraphy for northwestern Europe¹⁴ illustrated in Fig. 3. Palynology indicates an abrupt decrease in salinity concurrent with initiation of the excursion¹³.

Our results prompt a re-evaluation of the nature and origin of the Early Toarcian oceanic anoxic event (OAE). It could be argued that isotopically light carbon was derived from a huge deep-ocean reservoir which was receiving organic carbon from surface environments but not returning it to the surface as CO_2 until a time of sudden overturn, much as suggested for the negative carbon-isotope excursion at the Permian–Triassic boundary (for

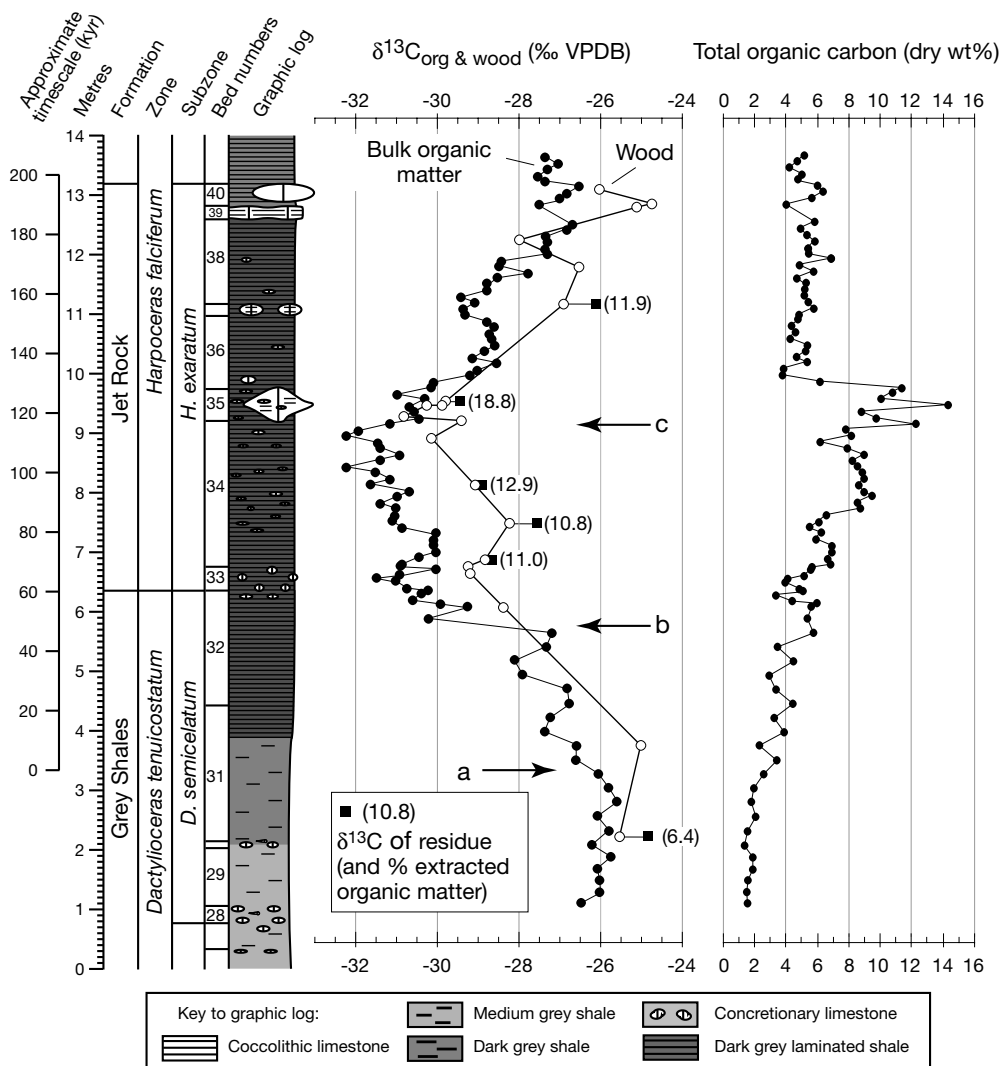


Figure 3 Carbon-isotope data through the lower Whitby Mudstone, Hawsker Bottoms, Yorkshire. Graphic log from ref. 30. Biostratigraphy from ref. 14 and references therein. Negative excursion begins at level a, shows a large abrupt decrease in $\delta^{13}\text{C}$ values at level b, and terminates at level c. No palaeontological or sedimentary evidence exists for a

hiatus at level b. Slight divergence between wood and bulk organic-isotope ratios may be due to changing relative abundance of organic components, but this cannot have been more than a minor factor in the genesis of the excursion. Approximate timescale is based on reported laminae thickness.

example, ref. 15). However, necessary consequences of this mechanism are deposition of widespread anoxic or strongly dysaerobic facies before the earliest *falciferum* zone, and coeval development of a positive carbon-isotope excursion in the shallow-oceanic and atmospheric reservoirs. Neither of these phenomena is known from the stratigraphic record¹.

Recent age determinations indicate coincidence of the Early Toarcian OAE with large igneous-province formation, as is also the case for other OAEs and the LPTM (for example, ref. 16). A U-Pb age of 181.4 ± 1.2 Myr BP from an ammonite-calibrated ash layer in western Canada corresponds to the *variabilis* zone of northwestern Europe¹⁷, and Ar-Ar dates from the Karoo–Ferrar continental flood basalts give ages of 183 ± 1 Myr BP (ref. 18). The Early Toarcian OAE is two ammonite zones below *variabilis*, and is therefore roughly 2 Myr older. Volcanogenic CO₂ from

Karoo–Ferrar sources ($\delta^{13}\text{C} \approx -7\text{‰}$) is unlikely to have effused rapidly enough to cause such a large excursion directly (compare with ref. 7). However, massive volcanism, together with intensified rift-related tectonic activity¹⁹, could conceivably have led to environmental changes sufficient to cause methane-hydrate dissociation⁷ in continental margin sediments, and generation of the negative $\delta^{13}\text{C}$ excursion ($\delta^{13}\text{C}_{\text{methane}} \approx -60\text{‰}$). If this explanation is correct, Early Toarcian methane release occurred during eustatic sea-level rise (estimated to be about 30–90 m over approximately 1.5 Myr (refs 10, 20)), and therefore increasing hydrostatic pressure and hydrate stability. Thus, the proximal trigger for hydrate dissociation must have been raised temperature at the ocean floor and consequent realignment of the temperature gradient in the fault-disrupted sediment pile.

An effective means to increase bottom-water temperatures would

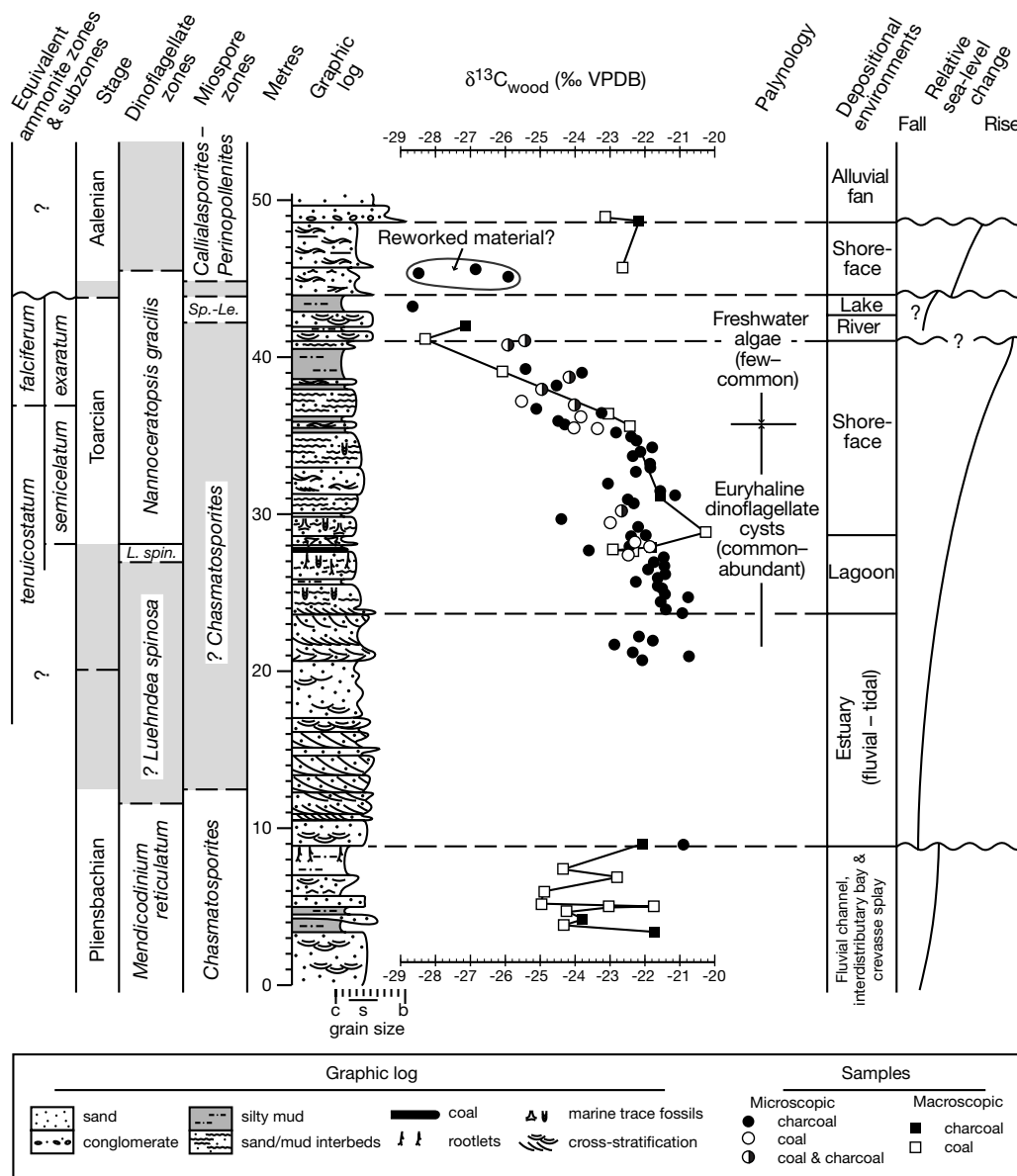


Figure 4 Stratigraphic and carbon-isotope data through the upper Bagå Formation at Korsodde, southwestern Bornholm. The up-section transition to marine deposits is gradual and precedes the initiation of the negative excursion, but the Early Toarcian strata are truncated by a sharp erosion surface at 44 m. Graphic log, palynology, microfossil zonation, inferred environments and sea-level changes are summarized from ref. 13 with

slight modification. Grey shading in biostratigraphic columns indicates uncertainty resulting from absence of index taxa. Stage assignment is based solely on microfossil data. Equivalent ammonite biostratigraphy is based on correlation of $\delta^{13}\text{C}$ excursion to the UK reference sections (Figs 2 and 3). *Sp.-Le.* indicates the *Spheripollenites*–*Leptolepides* zone. c, s and b are the clay, sand and boulder grain-size categories.

have been a change in the global thermohaline circulation such that bottom-water formation switched from high to low latitude. This general mechanism has been proposed to explain the release of methane during the LPTM. Physical palaeoceanographic modelling suggests that bottom-water temperatures in Tethys could have risen by up to 5 °C in this manner²¹. Water flow through the northwestern European seaway (Fig. 1) is likely to have been related to global thermohaline circulation (as in ref. 22): with southward flow driven by low-latitude deep-water formation, modelled mean salinity in the seaway is reduced by 3–5‰ compared to northward flow²¹. The data for northwestern Europe indicate that salinity reduction, initiation of black-shale deposition and the beginning of the negative carbon-isotope excursion are all synchronous. Although the OAE is undeniably a global phenomenon, the introduction of cool, nutrient-rich bottom waters of northerly origin could help to explain the extraordinary nature of the organic enrichment in the northwestern European seaway.

The negative excursion in organic matter is –4 to –7‰, depending on location and chosen baseline (Figs 3 and 4). Some of this variability may be attributed to taxonomic or taphonomic factors. The size of the negative excursion in marine carbonates is also variable (–2 to –5‰) with some considerable uncertainty stemming from diagenetic effects. Most geological data suggest a two-fold amplification of $\delta^{13}\text{C}_{\text{org}}$ variations relative to stratigraphically equivalent $\delta^{13}\text{C}_{\text{carb}}$ (ref. 16), thought to be a consequence of changing atmospheric $p\text{CO}_2$ occurring in parallel with reservoir-scale shifts in isotopic values. Although higher atmospheric CO_2 concentrations are a likely result of methane release²³ they are as yet unquantified by geological data. As a basis for discussion, we take the magnitude of the excursion in the total exchangeable carbon reservoir to be between –2 and –3.5‰.

The mass of methane-hydrate carbon necessary to cause the negative excursion over this short timescale can be estimated using simple mass-balance equations⁷. Taking present-day mass and $\delta^{13}\text{C}$ estimates, we calculate that 1.5×10^{18} to 2.7×10^{18} g of carbon is required for excursions of –2 or –3.5‰ respectively. These figures are 14–24% of the estimated present-day gas-hydrate reservoir (compare 14–19% for the LPTM using estimates of reservoir mass and isotopic composition derived from refs 7 and 24). If the synchronous burial of light organic carbon is taken into account, the mass of methane-derived carbon necessary to produce the excursion is very much larger. Release and oxidation of methane in such quantities would have reduced oceanic O_2 levels and thus helped promote organic-carbon burial, irrespective of any productivity changes driven by redistribution of nutrients. It is also likely that the bicarbonate and carbonate balance of the oceans would have been strongly disturbed and carbonate solubility increased⁷. In deep-water Tethyan settings, the common absence of primary carbonates carrying the negative carbon-isotope signal^{1,2} is thus a predicted consequence of gas-hydrate dissociation.

Our estimates indicate that the Early Toarcian OAE could have involved the release of twice as much methane carbon than was the case for the LPTM. A negative $\delta^{13}\text{C}$ excursion is also associated with initiation in the Cretaceous of the Aptian OAE, or Selli event (about 120 Myr ago): the isotopic shift in this case is –2.5 to –3‰ in carbonates²⁵ and as great as –7‰ in wood¹⁶. The Selli event, which shows many overall similarities to the Early Toarcian OAE, was also characterized by enhanced organic-carbon burial, and had an estimated duration of less than 400 kyr (ref. 26). Thus, the Early Toarcian OAE and Selli event²⁷ together appear to be the two largest proposed hydrate-dissociation events out of at least three similar occurrences over the past 200 Myr of Earth history. □

Received 14 January; accepted 25 May 2000.

1. Jenkyns, H. C. The early Toarcian (Jurassic) anoxic event: stratigraphic, sedimentary and geochemical evidence. *Am. J. Sci.* **288**, 101–151 (1988).

2. Jenkyns H. C. & Clayton, C. J. Lower Jurassic epicontinental carbonates and mudstones from England and Wales: chemostratigraphic signals and the early Toarcian anoxic event. *Sedimentology* **144**, 687–706 (1997).
3. Vakhrameev, V. A. *Jurassic and Cretaceous Floras and Climates of the Earth* (Cambridge Univ. Press, Cambridge, 1991).
4. Harries, P. J. & Little, C. T. S. The early Toarcian (Early Jurassic) and the Cenomanian–Turonian (Late Cretaceous) mass extinctions: similarities and contrasts. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **154**, 39–66 (1999).
5. Küspert, W. G. in *Cyclic and Event Stratification* (eds Einsele, G. & Seilacher, A.) 482–501 (Springer, Berlin, Germany, 1982).
6. Schouten, S., Van Kaam-Peters, H. M. E., Rijpstra, W. I. C., Schoell, M. & Sinninghe Damste, J. S. Effects of an oceanic anoxic event on the stable carbon isotopic composition of Early Toarcian carbon. *Am. J. Sci.* **300**, 1–22, (2000).
7. Dickens, G. R., O’Neil, J. R., Rea, D. K. & Owen, R. M. Dissociation of oceanic methane hydrate as a cause of the carbon isotope excursion at the end of the Paleocene. *Paleoceanography* **10**, 965–971 (1995).
8. Duarte, L. V. Clay minerals and geochemical evolution in the Toarcian-lower Aalenian of the Lusitanian basin (Portugal). *Cuad. Geol. Ibérica* **24**, 69–98 (1998).
9. Abbott, G. D., Ewbank, G., Edwards, D. & Guang-Yu Wang Molecular characterization of some enigmatic Lower Devonian fossils. *Geochim. Cosmochim. Acta* **62**, 1407–1418 (1998).
10. Hallam, A. Estimates of the amount and rate of sea-level change across the Rhaetian–Hettangian and Pliensbachian–Toarcian boundaries (latest Triassic to early Jurassic). *J. Geol. Soc. Lond.* **154**, 773–779 (1997).
11. Cope, J. C. W. Discussion on estimates of the amount and rate of sea-level change across the Rhaetian–Hettangian and Pliensbachian–Toarcian boundaries (latest Triassic to Early Jurassic). *J. Geol. Soc. Lond.* **155**, 421 (1998).
12. Norris, R. D. & Röhl, U. Carbon cycling and chronology of climate warming during the Palaeocene/Eocene transition. *Nature* **401**, 775–778 (1999).
13. Koppelhus, E. B. & Nielsen, L. H. Palynostratigraphy and palaeoenvironments of the lower to middle Jurassic Bagå Formation of Bornholm, Denmark. *Palynology* **18**, 139–194 (1994).
14. Howarth, M. K. The ammonite family Hildoceratidae in the Lower Jurassic of Britain. Part 1. *Palaeogr. Soc. Monogr.* **145**, 1–106 (1992).
15. Knoll, A. H., Bambach, R. K., Canfield, D. E. & Grotzinger, J. P. Comparative Earth history and Late Permian mass extinction. *Science* **272**, 452–457 (1996).
16. Gröcke, D. R., Hesselbo, S. P. & Jenkyns, H. C. Carbon-isotope composition of Lower Cretaceous fossil wood: Ocean–atmosphere chemistry and relation to sea-level change. *Geology* **27**, 155–158 (1999).
17. Pálfi, J., Parrish, R. R. & Smith, P. L. A U–Pb age from the Toarcian (Lower Jurassic) and its use for time scale calibration through error analysis of biochronologic dating. *Earth Planet. Sci. Lett.* **146**, 659–675 (1997).
18. Duncan, R. A., Hooper, P. R., Rehacek, J., Marsh, J. S. & Duncan, A. R. The timing and duration of the Karoo igneous event, southern Gondwana. *J. Geophys. Res.* **102**, 18127–18138 (1997).
19. Ziegler, P. A. *Geological Atlas of Western and Central Europe* (Shell Internationale Petroleum, The Hague, 1990).
20. Hesselbo, S. P. & Jenkyns, H. C. in *Mesozoic and Cenozoic Sequence Stratigraphy of European Basins* (eds de Graciansky, P. C. et al.) 560–581 (SEPM, Society for Sedimentary Geology, Tulsa, 1998).
21. Bjerrum, C. J. *Numerical Palaeoceanography Of A Jurassic Narrow Meridional Seaway: Transcontinental Currents And Global Ocean Feedbacks*. PhD thesis, Univ. Copenhagen (1999).
22. Shaffer, G. & Bendtsen, J. Role of the Bering Strait in controlling North Atlantic ocean circulation and climate. *Nature* **367**, 354–357 (1994).
23. Dickens, G. R. Methane oxidation during the Late Palaeocene thermal maximum. *Bull. Soc. Géol. Fr.* **171**, 37–49 (2000).
24. Kvenvolden, K. A. in *Gas Hydrates: Relevance to World Margin Stability and Climate Change* (eds Henriot, J.-P. & Mienert, J.) 9–30 (Geological Society, London, 1998).
25. Jenkyns, H. C. & Wilson, P. A. Stratigraphy, paleoceanography, and evolution of Cretaceous Pacific Guyots: relics from a Greenhouse Earth. *Am. J. Sci.* **299**, 341–392 (1999).
26. Fiet, N. Calibrage temporel de l’Aptien et des sous-étages associés par une approche cyclostratigraphique appliquée à la série pélagique de Marches-Ombrie (Italie centrale). *Bull. Soc. Géol. Fr.* **171**, 103–113 (2000).
27. Jahren, A. H. & Arens, N. C. Methane hydrate dissociation implicated in Aptian OAE events. *Geol. Soc. Am. Abs. Progs* **30**, 53 (1998).
28. Smith, A. G., Smith, D. G. & Funnell, B. M. *Atlas of Mesozoic and Cenozoic Coastlines* (Cambridge Univ. Press, Cambridge, UK, 1994).
29. Jiménez, A. P., Jiménez de Cisneros, C., Rivas, P. & Vera, J. A. The Early Toarcian anoxic event in the westernmost Tethys (Subbetic): Paleogeographic and paleobiogeographic significance. *J. Geol.* **104**, 399–416 (1996).
30. Hesselbo, S. P. & Jenkyns, H. C. in *Field Geology of the British Jurassic* (ed. Taylor, P. D.) 105–150 (Geological Society, London, 1995).

Acknowledgements

We thank F. Surlyk and G. Pedersen for help with this study. Useful critical comments were provided by J. Hudson. Palaeoceanographic modelling was carried out by C.J.B. in part while at the Geological Institute, University of Copenhagen. Organic carbon-isotope analyses were carried out at the University of Oxford Radiocarbon Unit thanks to T. O’Connell and M. Humm. D.M. Jones and P. Donohoe helped with Py-GC-MS analysis. Scientific discussion with L. Nielsen, G. Henderson and C. Jones is gratefully acknowledged. D.R.G. is funded by ESSO, C.J.B. by the Danish Natural Science Research Council, and H.S.M.B. by a NERC–industrial consortium (Rapid Global Geological Events project).

Correspondence and requests for materials should be addressed to S.P.H. (e-mail: stephen.hesselbo@earth.ox.ac.uk).

Evidence for a late chondritic veneer in the Earth's mantle from high-pressure partitioning of palladium and platinum

A. Holzheid^{*||}, P. Sylvester[†], H. St C. O'Neill[‡], D. C. Rubie[§] & H. Palme^{*}

^{*} Universität zu Köln, Institut für Mineralogie und Geochemie, Zùlpicherstrasse 49b, 50674 Köln, Germany

[†] Department of Earth Sciences, Memorial University of Newfoundland, 300 Prince Philip Drive, St John's NF A1B 3X4, Canada

[‡] Research School of Earth Sciences, Australian National University, Canberra 0200, Australia

[§] Bayerisches Geoinstitut, Universität Bayreuth, 95440 Bayreuth, Germany

The high-pressure solubility in silicate liquids of moderately siderophile 'iron-loving' elements (such as nickel and cobalt) has been used to suggest that, in the early Earth, an equilibrium between core-forming metals and the silicate mantle was established at the bottom of a magma ocean^{1,2}. But observed concentrations of the highly siderophile elements—such as the platinum-group elements platinum, palladium, rhenium, iridium, ruthenium and osmium—in the Earth's upper mantle can be explained by such a model only if their metal–silicate partition coefficients at high pressure are orders of magnitude lower than those determined experimentally at one atmosphere (refs 3–8). Here we present an experimental determination of the solubility of palladium and platinum in silicate melts as a function of pressure to 16 GPa (corresponding to about 500 km depth in the Earth). We find that both the palladium and platinum metal–silicate partition coefficients, derived from solubility, do not decrease with pressure—that is, palladium and platinum retain a strong preference for the metal phase even at high pressures. Consequently the observed abundances of palladium and platinum in the upper mantle seem to be best explained by a 'late veneer' addition of chondritic material to the upper mantle following the cessation of core formation.

The fundamental differentiation of the Earth was into an FeNi-rich metallic core and a surrounding silicate mantle. During this process, which is thought from isotopic evidence to have occurred soon after accretion (see, for example, ref. 9), the siderophile elements were preferentially partitioned into the core-forming metal, leaving the mantle depleted in these elements relative to their original solar or chondritic abundances. It has long been a goal of geochemistry that the study of the mantle abundances of these siderophile elements, together with knowledge of their metal–silicate partition coefficients as a function of physico-chemical variables such as temperature, pressure and oxygen fugacity, may shed light on how core formation occurred, and thus, possibly, also on how the Earth accreted. For example, can the mantle abundances of the siderophile elements be explained by metal/silicate partitioning starting with material of one composition, at a unique condition of temperature and pressure (implying that the Earth accreted homogeneously), or are more complex models needed? Recent studies of siderophile element partitioning between coexisting metal and silicate phases have led to two contrary hypotheses: (1) the present mantle signature of siderophile elements reflects equilibrium distribution between metal and peridotitic silicate liquid at the base of a deep (800–1,000 km, ~28 GPa) magma ocean during core formation^{1,2,10,11}. (2) On the basis of 1-atm

partition behaviour of siderophile elements, it has been suggested that the pattern of siderophile elements in the mantle implies heterogeneous accretion of the Earth, including a substantial component of oxidized material originating from greater heliocentric distances than the indigenous material at 1 astronomical unit^{9,12–14} (1 astronomical unit is the average Earth–Sun distance). As pointed out in ref. 15, there are not enough experimentally determined high-pressure partitioning data for all siderophile elements to test both models rigorously.

Of special interest in this context are the highly siderophile elements (HSEs, defined as having metal/silicate distribution coefficients $D^{\text{met/sil}} > 10^4$ [$D^{\text{met/sil}}$ = concentration (by wt) of element of interest in the metal phase/concentration (by wt) of element of interest in the silicate phase], and consisting of the six platinum group elements (Ru, Rh, Pd, Os, Ir and Pt) plus Re and Au. Because of the high values of $D^{\text{met/sil}}$ (1-atm data, refs 3–8), core formation should have stripped the silicate mantle of these elements to vanishingly small levels ($< 10^{-4}$ of their chondritic abundances). Yet these elements are only depleted in the upper mantle by a factor of 150 relative to C1 chondrites^{16,17}. The degree of depletion of all eight HSEs appears to be similar^{17,18}, despite their 1-atm $D^{\text{met/sil}}$ varying by several orders of magnitude^{3–8}. Both the overabundance of the HSEs and their near-chondritic relative abundances in the Earth's mantle would appear to be best explained by the addition of about 0.7% of the Earth's total mass of chondritic material to the Earth's mantle, after the core had formed. This is known as the 'late veneer' hypothesis, first proposed by Kimura *et al.* (ref. 19).

There is at present, however, almost no information on how the values of $D^{\text{met/sil}}$ for the HSEs vary with pressure. To address this question we have experimentally determined $D^{\text{met/sil}}$ for Pd and Pt as a function of pressure. Palladium was chosen for this work because in 1-atm experiments this element, unlike other HSEs, did not form tiny micronuggets in the silicate^{3–6}; such micronuggets greatly complicate the interpretation of metal/silicate partitioning relationships. Metal/silicate equilibrium distribution of Pd during core formation would require a high-pressure/high-temperature $D^{\text{met/sil}}$ of about 1,000, which is about 1,000 times lower than the experimentally derived 1-atm $D^{\text{met/sil}}$ value of 10^6 (ref. 3) at oxygen fugacities relevant for core formation (IW–2.3, that is, 2.3 log units below the iron–wüstite buffer) and 1,500 °C. This required decrease in the metal/silicate partition coefficients of Pd should be quite apparent in experiments. Despite problems with nugget formation at 1 atm and reducing conditions^{4,5} we have also attempted to determine $D^{\text{met/sil}}$ for Pt as a function of pressure. The value of $D^{\text{met/sil}}$ for platinum ($D_{\text{Pt}}^{\text{met/sil}}$) at one atmosphere is at least three orders of magnitude greater than $D_{\text{Pd}}^{\text{met/sil}}$ at comparable temperature and pressure. Thus, if the mantle abundance of Pt is to be explained by equilibrium partitioning processes, the effect of pressure on $D_{\text{Pt}}^{\text{met/sil}}$ would have to be even more pronounced.

The experiments were performed at pressures between 0.5 and 16 GPa (corresponding to a depth of about 500 km in the Earth) and temperatures of 1,100 to 1,600 °C in piston–cylinder and multi-anvil devices at the Bayerisches Geoinstitut, and in piston–cylinder devices at the Australian National University. To determine the temperature dependence of Pd solubility at high pressure, piston–cylinder experiments were performed at a constant pressure of 1.5 GPa and temperatures between 1,100 and 1,400 °C. Experimental parameters and results are given in Table 1. The effective oxygen fugacity that had prevailed during the experiment was calculated relative to the IW (Fe–FeO) buffer from the activities of Fe and FeO in metal and silicate phases, assuming negligible excess molar volumes of Fe in the Fe–Pd and Fe–Pt alloys and FeO in the silicate liquid:

$$\log f_{\text{O}_2}^{\text{exp}} = \text{IW} - 2 \log \frac{a_{\text{FeO}}}{a_{\text{Fe}}} = \text{IW} - 2 \log \frac{X_{\text{FeO}} \gamma_{\text{FeO}}}{X_{\text{Fe}} \gamma_{\text{Fe}}}$$

Here $\gamma_{\text{FeO}} \approx 1.7$ (ref. 20), and $\gamma_{\text{Fe}} = 0.010$ – 0.015 (ref. 21).

^{||} Present address: Westfälische Wilhelms-Universität Münster, Institut für Mineralogie, Corrensstraße 24, 48149 Münster, Germany.

The calculated oxygen fugacities in the high-pressure experiments range from 1.2 to 4.4 log units above the IW buffer. Values of $D^{\text{met/sil}}$ for Pd and Pt were calculated using the concentrations of Pd and Pt in the metal and silicate phase (Table 1).

In order to compare partitioning data from different temperatures, oxygen fugacities and metal compositions, the data were recalculated to a single temperature of 1,500 °C, an oxygen fugacity of 2.3 log units below the IW buffer (the appropriate oxygen fugacity for equilibrium between the Fe-rich metal in the core with 80 wt% Fe, and a silicate mantle with 8 wt% FeO, ref. 9), assuming Pd^+ and Pt^{2+} as the dominant species in the silicate melt^{3–5}, and to infinite dilution of Pd and Pt in solid Fe metal. The correction for temperature was made using the experimentally determined temperature dependence of $D_{\text{Pd}}^{\text{met/sil}}$ from this study and ref. 3, and of $D_{\text{Pt}}^{\text{met/sil}}$ from ref. 4. The correction to pure Fe metal was made using the activity–composition relations of Fe–Pd alloys and of Fe–Pt alloys of ref. 21. The recalculated ratios are given in Table 1 and plotted in Fig. 1a as function of pressure. The symbols represent average $D^{\text{met/sil}}$ values based on up to four individual spot analyses of the silicate phase. The error bars reflect the scatter of the individual analyses of the silicates, that is, the lowest and highest HSE contents. The range of measured HSE contents in the silicate are, in many cases, significantly larger than the statistical error of individual analyses (< 15%), and are probably due to the formation of nuggets of Pd and Pt. The calculated partition coefficients may therefore represent minimum values of $D^{\text{met/sil}}$. Larger real values of $D_{\text{Pd}}^{\text{met/sil}}$ and $D_{\text{Pt}}^{\text{met/sil}}$ would of course only strengthen our case, that

$D_{\text{Pd}}^{\text{met/sil}}$ and $D_{\text{Pt}}^{\text{met/sil}}$ do not decrease with pressure.

There are some differences between the 1-atm experiments performed here ($D_{\text{Pd}}^{\text{met/sil}}$, filled squares; $D_{\text{Pt}}^{\text{met/sil}}$, grey squares; Fig. 1a) and those of refs 3–5 (filled and grey-shaded vertical boxes), recalculated to 1,500 °C and IW–2.3. This may be ascribed to differences of the Pd_2O and PtO activity coefficients in Fe-free systems versus Fe-containing systems (anorthite–diopside eutectic silicate in refs 3–5 versus high MgO–basalt here, that is, FeO-free silicate versus FeO-containing silicate). Similar values of $D_{\text{Pd}}^{\text{met/sil}}$ and $D_{\text{Pt}}^{\text{met/sil}}$ derived from HSE solubilities in extremely low-melting FeO-containing K–Na-rich silicates and FeO-containing MgO-rich silicates indicate only slight dependences of the Pd_2O and PtO activity coefficients on melt composition in Fe-containing systems. Also, Pt solubilities in FeO-free silicate melts show only minor dependence on the composition of the melt within a large compositional range (ref. 4). The high-pressure data cover a relatively broad band that overlaps with the 1-atm data of our study. The important conclusion is that there is no evidence that $D^{\text{met/sil}}$ values for Pd and Pt decrease with increasing pressure: at pressures of 0.5–16 GPa, $D^{\text{met/sil}}$ values are of the order of 10^5 to 10^6 (Pd) and $> 10^7$ to 10^9 (Pt).

The present experiments were between metallic Pd (with minor Fe) and silicate, but they can readily be applied to FeNi alloys which are thought to be the major constituent of the Earth's core. The activity coefficient of Pd in FePd alloys at nominally 0% Pd is 0.15 for solid Fe, and for liquid Fe is 2.95 at 1,809 K and 1.46 at 3,500 K (ref. 3). In extrapolations to high-pressure/high-temperature conditions liquid iron would have to be considered. This would reduce

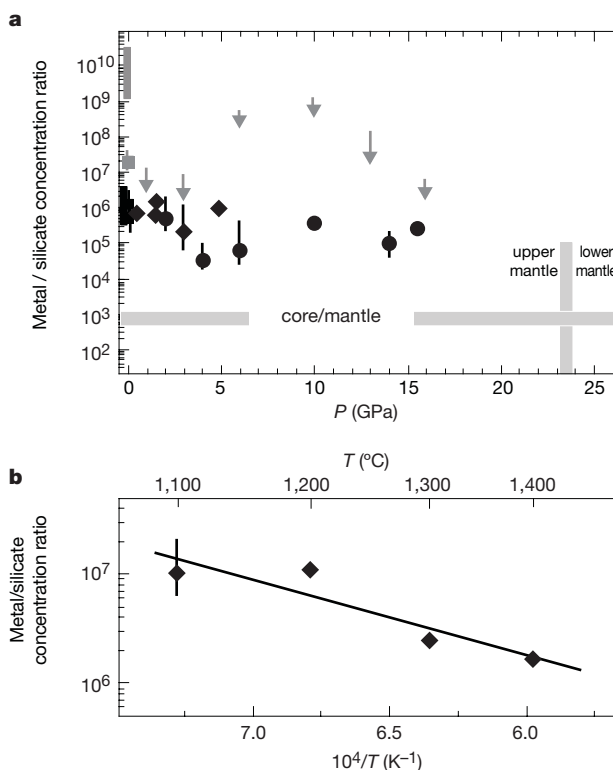


Figure 1 Pd and Pt metal/silicate concentration ratios. **a**, Pd and Pt metal/silicate concentration ratios plotted as function of total pressure recalculated to 1,500 °C, 2.3 log units below the Fe–FeO equilibrium (IW–2.3), and infinite dilution in solid Fe metal. Experiments with low-melting K_2O – Na_2O –FeO–silicates and $\text{Pd}_{90}\text{Fe}_{10}$ alloys are symbolized by filled circles (BGI) and diamonds (ANU). Experiments with low-melting K_2O – Na_2O –FeO–silicates and $\text{Pt}_{90}\text{Fe}_{10}$ alloys are symbolized by grey triangles (BGI). Experiments at one atmosphere with MgO-rich basalt as silicate starting material are symbolized by filled squares (Pd) and grey squares ($\text{Pt}_{90}\text{Fe}_{10}$). Experiments at one atmosphere with anorthite–diopside eutectic silicate as silicate starting material (refs 3–5; recalculated to 1,500 °C, IW–2.3) are symbolized by filled (Pd) and grey vertical boxes

(Pt). All symbols are average $D^{\text{met/sil}}$ values, the error bars cover the range in HSE content based on up to four individual analyses. The horizontal solid line represents the present Earth's core–mantle concentration ratio of HSE, the vertical line the upper mantle–lower mantle boundary. Recalculated $D^{\text{met/sil}}$ values differ from today's core–mantle ratio by 2 to 3 orders of magnitude (Pd) and by more than 4 orders of magnitude (Pt), respectively. Similar $D^{\text{met/sil}}$ values for HSEs and core–mantle ratio would be required to explain the HSE content of the Earth's mantle by a simple metal/silicate (core–mantle) equilibrium distribution. **b**, Dependence on temperature of metal/silicate concentration ratios of Pd at a constant pressure of 1.5 GPa and recalculated to IW–2.3.

Table 1 Distribution of Pd and Pt between Pd₉₀Fe₁₀ and Pt₉₀Fe₁₀ alloy and silicates

No.	<i>T</i> (°C)	<i>P</i> (GPa)	Time (h)	ΔW	Metal phase			Silicate phase			$D^{\text{met/sil}}$ (1)	$D^{\text{met/sil}}$ (2)
					Fe (wt %)	Pd (wt %)	Pt (wt %)	FeO (wt %)	Pd (p.p.m.)	Pt (p.p.m.)		
High-pressure experiments (silicate: KNSF silicate; metal–Pd ₉₀ Fe ₁₀ alloy, Pt ₉₀ Fe ₁₀ alloy)												
Piston-cylinder experiments												
C-997	1,300	0.5	24	2.3	9.52(8)	90.7(5)		2.27(5)	4.42		2.1×10 ⁵	6.7×10 ⁵
C-1000	1,100	1.5	36	3.8	8.16(35)	91.4(12)		8.23(7)	2.93		3.1×10 ⁵	5.9×10 ⁵
C-973	1,200	1.5	24	4.0	7.94(9)	92.7(13)		8.78(9)	3.08		3.0×10 ⁵	1.4×10 ⁶
C-977	1,300	1.5	8	4.0	7.70(18)	92.5(4)		7.77(3)	13.8		6.7×10 ⁴	6.2×10 ⁵
C-972	1,400	1.5	2	3.4	10.9(6)	89.0(3)		11.3(3)	14.0		6.4×10 ⁴	6.5×10 ⁵
PC-18	1,250	2	20	3.0	8.82(12)	91.5(7)		3.92(11)	7.08		1.3×10 ⁵	4.8×10 ⁵
C-999	1,300	3	24	4.4	6.66(21)	93.5(4)		7.96(17)	57.3		1.6×10 ⁴	2.1×10 ⁵
B-39	1,300	4.9	43	3.2	10.4(1)	91.0(6)		8.05(12)	5.76		1.6×10 ⁵	9.5×10 ⁵
PC-17	1,400	1	20	1.9	10.6(1)		88.5(12)	2.89(7)		16.2	5.5×10 ⁴	4.1×10 ⁶
PC-16	1,400	3	21	1.2	10.5(2)		88.7(14)	1.33(5)		11.1	8.0×10 ⁴	2.7×10 ⁶
Multi-anvil experiments												
V-29	1,300	4	20	2.2	12.8(2)	87.7(9)		6.39(28)	68.4		1.3×10 ⁴	3.3×10 ⁴
H-937	1,300	6	19	3.3	9.24(7)	91.2(10)		5.40(18)	87.0		1.0×10 ⁴	6.1×10 ⁴
V-30	1480	10	24	2.8	11.5(11)	89.5(10)		6.35(47)	27.2		3.3×10 ⁴	3.6×10 ⁵
H-950	1,500	14	12	3.4	8.51(80)	91.7(14)		6.62(47)	189		4.9×10 ³	9.9×10 ⁴
H-1184	1,550	15.5	25	3.9	5.77(20)	94.9(10)		2.29(27)	153		6.2×10 ³	2.5×10 ⁵
H-934	1,550	6	20	2.2	10.3(2)		89.4(7)	4.68(13)		0.34	2.6×10 ⁶	2.9×10 ⁸
H-943	1,550	10	14	2.1	10.4(2)		89.1(13)	4.14(15)		0.14	6.4×10 ⁶	6.1×10 ⁸
H-1181	1,600	13	7	2.8	11.0(3)		88.8(8)	9.37(16)		7.64	1.2×10 ⁵	2.7×10 ⁷
1973	1,550	16	24	2.2	10.1(1)		89.2(28)	4.66(11)		38.8	2.3×10 ⁴	2.8×10 ⁶
1-atm experiments (silicate: BK silicate; metal–pure Pd, Pt ₉₀ Fe ₁₀ alloy)												
PdBK-1	1,454	10 ^{−4}	53	2.9	6.87(76)	92.9(7)		1.87(12)	9.32		1.0×10 ⁵	1.3×10 ⁶
PdBK-2	1,454	10 ^{−4}	5	3.3	5.68(14)	93.4(7)		4.44(7)	8.60		1.1×10 ⁵	5.2×10 ⁵
PtFe-1	1,300	10 ^{−4}	24	−0.6	9.87(7)		91.2(1)	0.182(46)		0.22	4.1×10 ⁶	1.9×10 ⁷

Metals and major-element concentrations of silicates are analysed by electron microprobe; listed Pd and Pt contents in silicates are average values of up to four individual laser ablation ICP-MS analyses. Numbers in parentheses are 2σ standard deviations. The entry 9.52(8) should be read 9.52 wt% $\pm$ 0.08 wt%; standard deviations of individual Pd and Pt analyses are in most cases less than 15% (relative). All high-pressure experiments were performed at BGI, except experiments labelled B and C that were done at ANU. For comparison with 1-atm distribution behaviour experiments under controlled temperature and oxygen fugacity with FeO-containing 'BK' silicate but pure Pd and Pt₉₀Fe alloy as metal phase are performed using 1-atm furnaces at the University of Cologne (PdBK-1, -2; PtFe-1). HSE-concentrations of these silicates are analysed by instrumental neutron activation analysis. $D^{\text{met/sil}}$ (1) indicates metal/silicate distribution coefficients at run conditions; $D^{\text{met/sil}}$ (2) indicates metal/silicate distribution coefficients recalculated to 1,500 °C, ΔW = 2.3, and infinite dilution in solid Fe metal (plotted in Fig. 1a).

the $D^{\text{met/sil}}$ by, at most, a factor of 1.5, as the metal/silicate partition coefficient for Pd decreases with increasing temperature. A similar effect would be expected for $D^{\text{met/sil}}$ values.

If core–mantle equilibrium established the HSE contents of the Earth's mantle, we would expect an effective $D^{\text{met/sil}}$ for HSEs of about 1,000 (indicated as horizontal bar in Fig. 1a) at the pressure and temperature where metal–silicate equilibration occurred. The present experiments have shown that there is no tendency for the metal/silicate partition coefficients of Pd or Pt to decrease with pressure. A slight increase in the solubility of Pd in silicate liquid with increasing temperature was found in ref. 3 and in this study at 1.5 GPa (Fig. 1b). Extrapolation of this trend to high temperatures reduces the $D^{\text{met/sil}}$ from $\sim 4 \times 10^6$ at 1,350 °C to $\sim 8 \times 10^3$ at 3,000 °C at one bar (ref. 3) and to $\sim 2 \times 10^4$ at 1.5 GPa. This is still significantly above the value of about 1,000 required to explain the present upper-mantle Pd content by metal–silicate equilibration. At 2,100–2,400 °C, the core–mantle equilibration temperature assumed in ref. 1, the difference between calculated and expected partitioning would be about two orders of magnitude. The Pt solubility also increases slightly with increasing temperature; furthermore, the temperature dependence seems to correlate with the absolute oxygen fugacity value⁴. The reason for that behaviour is unclear. However, the increase of Pt solubility, that is, the decrease of the $D^{\text{met/sil}}$ at higher temperatures, is too small to explain the observed Pt abundances in the Earth's upper mantle. In addition, equilibrium partitioning would imply exactly the same temperature dependence of the $D^{\text{met/sil}}$ and $D^{\text{met/sil}}$ to satisfy the chondritic mantle ratio of the two metals.

Neither temperature nor, as we show here, pressure, seem able to decrease the values of $D^{\text{met/sil}}$ or $D^{\text{met/sil}}$ sufficiently to explain their unexpected high abundances in the Earth's mantle. According to ref. 1, sulphur dissolved in the metal phase does not have a significant influence on the metal/silicate partition coefficients of Ni or Co at high pressure. Although the influence of light

elements—for example, sulphur, carbon or hydrogen—on the solubility of HSEs in the coexisting silicate phase at elevated pressure is not known, it seems unlikely that such light elements as might be present in core-forming metal could account for the difference between calculated and expected partitioning. In addition, there is other evidence for the late veneer model. Ratios among HSEs in the Earth's mantle are essentially chondritic^{17,18}, and, to explain this, equilibrium models require nearly identical values of $D^{\text{met/sil}}$ for all HSEs. In contrast, at one atmosphere, constant temperature (1,500 °C) and oxygen fugacities relevant to core formation, $D^{\text{met/sil}}$ values for HSEs differ by several orders of magnitude (for example, $D^{\text{met/sil}}_{\text{Pd}} = 1.1 \times 10^6$ (ref. 3), $D^{\text{met/sil}}_{\text{Pt}} = 9.8 \times 10^9$ (ref. 4), $D^{\text{met/sil}}_{\text{Ir}} \approx 10^{12}$ (ref. 6)). The combined effects of pressure and temperature would have to eliminate these differences and produce identical partition coefficients. Another argument for the lack of equilibration between mantle and core is the near chondritic Re/Os ratio (to within 7%) in the primitive upper mantle of the Earth, as inferred from Os isotopes²². Core formation, that is, metal–silicate separation, would have fractionated Re from Os severely, so addition of a late chondritic veneer to the Earth, after core formation was completed, appears necessary.

The data presented here argue against equilibrium distribution of HSEs between core and mantle. They cannot, however, be used for or against homogeneous accretion models involving major fractions of the Earth. About 0.7% of the total mass of the Earth as a chondritic component is enough to provide the mantle inventory of HSEs. The late veneer is only 'visible' through the abundances of HSEs, as its contribution to other elements (even Ni and Co) is negligible. □

Methods

Experimental and analytical procedures

In the experiments at the Bayerisches Geoinstitut (BGI), commercially available Pd₉₀Fe₁₀ and Pt₉₀Fe₁₀ alloys (Goodfellow) were fabricated into capsules and equilibrated with an

FeO-containing K-Na-rich silicate ('KNSF': SiO₂ 64.3 wt%, K₂O 18.6 wt%, Na₂O 12.4 wt%, FeO 4.6 wt%) with a low melting temperature. The experiments at the Australian National University (ANU) used Pd₉₀Fe₁₀ alloys made by melting mixes of Pd and Fe powders. These alloys were also fabricated into capsules and equilibrated with the FeO-containing K-Na-rich silicate. The silicate material was prepared from a mixture of reagent grade oxides and carbonates (K, Na), melted under air, and quenched to glass. Small cylinders were drilled out of the KNSF silicate glass and were inserted into capsules fabricated from the Pd₉₀Fe₁₀ and Pt₉₀Fe₁₀ alloys (BGI experiments). In the case of ANU experiments, silicate glass powders were loaded into the capsules. The metal phases of interest were used as capsule material to avoid contamination of the sample with capsule material in the metal-silicate partition experiments; for example, contamination by MgO or Al₂O₃ of silicate phases by using MgO- or Al₂O₃-capsules as shown in, for example, ref. 23. Graphite was used as the heater in piston-cylinder furnace assemblies and LaCrO₃ as the heater in multi-anvil furnaces. The entire assemblies were placed in a vacuum oven at 250 °C overnight before the experiment. Details of the experimental design and method can be found in refs 24 (piston cylinder) and 25 (multi-anvil). After reaching the desired pressure, samples were heated to the desired temperatures (1,100–1,600 °C). Experimental conditions were chosen so that solid metal and liquid silicate coexisted during the experiments. The low-melting KNSF-silicate remained above its liquidus within the investigated *P*–*T*-range. Run durations were up to 36 h, and the experiments were terminated by turning off the power to the heater. The temperature drop was sufficiently rapid to quench the silicate liquids into glasses. Metal compositions and major-element silicate compositions were determined by electron microprobe (Cameca Camebax, University of Cologne; 10 nA, 20 kV, counting time: 40 s). Palladium and Pt concentrations in silicates were analysed by an ultraviolet (quadrupled Nd-YAG) laser ablation-PlasmaQuad PQ-2+ inductively coupled plasma-mass spectrometer (ICP-MS) system at Memorial University, Newfoundland, Canada, using procedures described in ref. 26. The beam spot was 100 µm diameter, laser repetition rate 10 Hz, laser energy density 40 J cm⁻², and counting time 60 s. The Pd and Pt standards were homogeneous synthetic silicate glasses whose Pd and Pt concentrations had been measured previously by instrumental neutron activation analysis. The Si content of each individual silicate sample was used as an internal standard for the laser ablation ICP-MS analyses. To enhance the comparison between the high-pressure results and 1-atm data, additional 1-atm experiments with pure Pd and Pt₉₀Fe₁₀-metal and an FeO-containing melt ('BK': SiO₂ 49.1 wt%, CaO 19.2 wt%, MgO 10.6 wt%, Al₂O₃ 14.1 wt%, FeO 7.0 wt%) as silicate starting material were performed under controlled temperature and oxygen fugacity in gas mixing furnaces at the University of Cologne using the loop technique as described in ref. 27. The Pd and Pt contents in these silicate samples were analysed by instrumental neutron activation analysis at the University of Cologne.

Received 23 December 1999; accepted 6 June 2000.

- Li, J. & Agee, C. B. Geochemistry of mantle-core differentiation at high pressure. *Nature* **381**, 686–689 (1996).
- Righter, K. & Drake, M. J. Metal-silicate equilibrium in a homogeneously accreting Earth: new results for Re. *Earth Planet. Sci. Lett.* **146**, 541–553 (1997).
- Borisov, A., Palme, H. & Spettel, B. Solubility of palladium in silicate melts: implications for core formation in the Earth. *Geochim. Cosmochim. Acta* **58**, 705–716 (1994).
- Borisov, A. & Palme, H. Experimental determination of the solubility of platinum in silicate melts. *Geochim. Cosmochim. Acta* **61**, 4349–4357 (1997).
- Ertel, W. *et al.* Solubilities of platinum and rhodium in a haplobasaltic melt at 1300 °C. *Geochim. Cosmochim. Acta* **63**, 2439–2449 (1999).
- Borisov, A. & Palme, H. The solubility of iridium in silicate melts: new data from experiments with Ir₁₀Pt₉₀ alloys. *Geochim. Cosmochim. Acta* **59**, 481–485 (1995).
- Borisov, A. & Palme, H. Noble metal/silicate melt equilibria: experimental investigations and geochemical implications. *Eur. J. Mineral.* **9** (Suppl.), 59 (1997).
- O'Neill, H. St C. *et al.* Experimental petrochemistry of some highly siderophile elements at high temperatures, and some implications for core formation and the mantle's early history. *Chem. Geol.* **120**, 255–273 (1995).
- O'Neill, H. St C. & Palme, H. In *The Earth's Mantle, Composition, Structure and Evolution* (ed. Jackson, I.), 3–126 (Cambridge Univ. Press, Cambridge, 1998).
- Murthy, V. R. Early differentiation of the Earth and the problem of mantle siderophile elements: a new approach. *Science* **253**, 303–306 (1991); corrections **253**, 1467 (1991).
- Walter, M. Screws tighten on the core. *Nature* **398**, 563–565 (1999).
- Ringwood, A. E. The Earth's core: its composition, formation and bearing upon the origin of the Earth. *Proc. R. Soc. Lond. A* **395**, 1–46 (1984).
- Wänke, H., Dreibus, G. & Jagoutz, E. In *Archean Geochemistry* (ed. Kröner, A.) 1–24 (Springer, Berlin, 1984).
- O'Neill, H. St C. The origin of the Moon and the early history of the Earth—a chemical model. Part 2: The Earth. *Geochim. Cosmochim. Acta* **55**, 1159–1172 (1991).
- Walter, M. At the magma ocean floor. *Nature* **381**, 646–647 (1996).
- Morgan, J. W., Wandless, G. A., Petrie, R. K. & Irving, A. J. Composition of the Earth's upper mantle-I. Siderophile trace elements in ultramafic nodules. *Tectonophysics* **75**, 47–67 (1981).
- Jagoutz, E. *et al.* The abundances of major, minor and trace elements in the Earth's mantle as derived from primitive ultramafic nodules. *Proc. Lunar Planet. Sci. Conf.* **10**, 2031–2050 (1979).
- Morgan, J. W. Ultramafic xenoliths: clues to Earth's late accretionary history. *J. Geophys. Res.* **91**, 12375–12387 (1986).
- Kimura, K., Lewis, R. S. & Anders, E. Distribution of gold and rhenium between nickel-iron and silicate melts: implications for the abundances of siderophile elements on the Earth and Moon. *Geochim. Cosmochim. Acta* **38**, 683–701 (1974).
- Holzheid, A., Palme, H. & Chakraborty, S. The activities of NiO, CoO and FeO in silicate melts. *Chem. Geol.* **139**, 21–38 (1997).
- Hultgren, R., Desai, P. D., Hawkins, D. T., Gleiser, M. & Kelley, K. K. *Selected Values of the Thermodynamic Properties of Binary Alloys* (American Society for Metals, Metals Park, Ohio, 1973).

- Meisel, T., Walker, R. J. & Morgan, J. W. The osmium isotopic composition of the Earth's primitive upper mantle. *Nature* **383**, 517–520 (1996).
- Hillgren, V. J., Drake, M. J. & Rubie, D. C. High-pressure and high-temperature experiments on core-mantle segregation in the accreting Earth. *Science* **264**, 1442–1445 (1994).
- Seckendorff, V. & O'Neill, H. St C. An experimental study of Fe-Mg partitioning between olivine and orthopyroxene at 1173, 1273, and 1423 K and 1.6 GPa. *Contrib. Mineral. Petrol.* **113**, 196–207 (1993).
- Rubie, D. C., Karato, S., Yan, H. & O'Neill, H. St C. Low differential stress and controlled chemical environment in multi-anvil high pressure experiments. *Phys. Chem. Miner.* **20**, 315–322 (1993).
- Sylvester, P. J. & Eggins, S. M. Analysis of Re, Au, Pt, Pd, and Rh in NIST glass certified reference materials and natural basalt glasses by laser ablation-ICP-MS. *Geostand. News* **21**, 215–229 (1997).
- Holzheid, A. & Palme, H. The influence of FeO on the solubilities of Co and Ni in silicate melts. *Geochim. Cosmochim. Acta* **60**, 1181–1193 (1996).

Acknowledgements

High pressure experiments were performed at the Bayerisches Geoinstitut under the EU 'TMR - Large Scale Facilities' (D.C.R.) programme.

Correspondence should be addressed to A.H. (e-mail: holzheid@uni-muenster.de).

Mutation and sex in a competitive world

Joel R. Peck & David Waxman

Centre for the Study of Evolution, The University of Sussex, Brighton BN1 9QG, UK

How do deleterious mutations interact to affect fitness? The answer to this question has substantial implications for a variety of important problems in population biology, including the evolution of sex^{1–3}, the rate of adaptation^{4,5} and the conservation of small populations^{3,6–8}. Here we analyse a mathematical model of competition for food in which deleterious mutations affect competitive ability. We show that, if individuals usually compete in small groups, then competition can easily lead to a type of genetic interaction known as synergistic epistasis. This means that a deleterious mutation is most damaging in a genome that already has many other deleterious mutations. We also show that competition in small groups can produce a large advantage for sexual populations, both in mean fitness and in ability to resist invasion by asexual lineages. One implication of our findings is that experimental efforts to demonstrate synergistic epistasis may not succeed unless the experiments are redesigned to make them much more naturalistic.

Consider a population of organisms for which generations are discrete, so that parents die at around the time that their offspring are produced. Let us define an adult as an individual that survives to reproductive age. We define the fitness associated with a particular genotype as the mean number of surviving offspring produced by adults with that genotype. In other words, when calculating fitness, we only count offspring that survive to adulthood.

Let us suppose that the fitness associated with a genotype depends only on the number of deleterious mutations. It is useful to normalize fitness, so we define a quantity, w_k , which is the fitness associated with having k deleterious mutations divided by the fitness associated with having zero deleterious mutations (thus, $w_0 = 1$). That is, w_k is the relative fitness of adults with k mutations.

If mutations act entirely independently of each other, then a plot of the logarithm of w_k as a function of k (the number of mutations) produces a straight line with a negative slope. This is the case where epistasis is absent, and mutations combine to affect fitness in a multiplicative fashion. If each mutation added to the genome has a greater deleterious effect than preceding mutations, then we have synergistic epistasis, leading to a curve where the rate of decrease in the logarithm of w_k increases with k (Fig. 1a). If each mutation

added to the genome has a smaller deleterious effect on fitness than preceding mutations, then we have diminishing returns epistasis, where the rate of decrease in the logarithm of w_k declines with k (Fig. 1b).

Let $\bar{w}$ represent the mean value of w_k at equilibrium (that is, $\bar{w}$ is the mean relative fitness). Analysis of simple models shows that, in the case of no epistasis, we have $\bar{w} = e^{-U}$, where U is the expected number of new mutations per offspring^{1,9}. For diminishing-returns epistasis these models yield $\bar{w} < e^{-U}$, and for synergistic epistasis, $\bar{w} > e^{-U}$.

The theoretical findings have focused the attention of experimentalists on synergistic epistasis^{1,10–16}. However, there is no strong experimental support for the idea that synergistic epistasis is ubiquitous^{13,16}. This is surprising because, in the absence of synergistic epistasis, it is difficult to explain how species with relatively small local populations (for example, mammals and trees) can survive genetic drift^{3,7}. Populations with high genomic mutation rates (U) are also difficult to explain without synergy, and there is evidence that high values of U may be common^{3,17–21}.

The present state of nature is puzzling, given the lack of success in demonstrating synergistic epistasis. We believe that the explanation is that synergy emerges in nature primarily from competition for resources among small groups of genetically diverse individuals. This type of situation is common, but it is more complex than what has been considered in most of the experimental work on synergy^{1,10–16}. Several authors have speculated that synergy emerges from competition for resources^{22,23}. However, as we shall see, there is no general reason to expect competition for resources to lead to synergy unless competition generally takes place in small groups. The authors who have gone furthest to emphasize the importance of competition in small groups as a source of synergy are A. S. Kondrashov¹ and W. D. Hamilton *et al.*²⁴, but they discussed small-group competition very briefly and without analysis of a formal model.

To understand the importance of competition, let us consider a simple model. We will focus on one specific case, but we believe that analysis of this case provides general insight.

Consider a diploid and outcrossing organism that depends on consuming a resource called ‘food’. Generations are discrete, and the expected number of offspring for a female adult is proportional to the amount of food she eats between reproductive maturity and death. Offspring are male or female, with equal probability. For convenience, we measure food in units such that the expected number of offspring increases by one for each unit of food consumed. We also make the convenient assumption that food appears in discrete quantities (these might be seeds, fruits or individual prey animals).

Each offspring is produced by a female adult after mating with a male adult, and the probability that a male adult will contribute to a particular offspring is assumed to be proportional to the amount of food that the male has consumed since maturation. This may reflect a better ability to compete for mates among better-fed males. Alternatively, for a ‘broadcast breeder’, this may reflect proportionality between the amount of food consumed and the number of male gametes produced.

Standard mendelian segregation and free recombination are involved in the production of gametes. A very large number of loci are susceptible to deleterious mutations. During gamete production, mutations occur independently at randomly selected loci, and the expected number of new mutations per gamete is $U/2$. Thus, the expected number of new mutations per offspring is U .

We assume that an adult’s genetic constitution controls the rate at which it can consume food. An adult with k mutations can consume food at a rate of C_k (where $C_k > 0$). The mutations are deleterious, and thus we assume that $C_{k+1}/C_k \leq (1 - \epsilon)$, for some ϵ that satisfies $0 < \epsilon < 1$. Note that these assumptions are fairly general in that many different relationships between the number of mutations and competitive ability are allowed within the model.

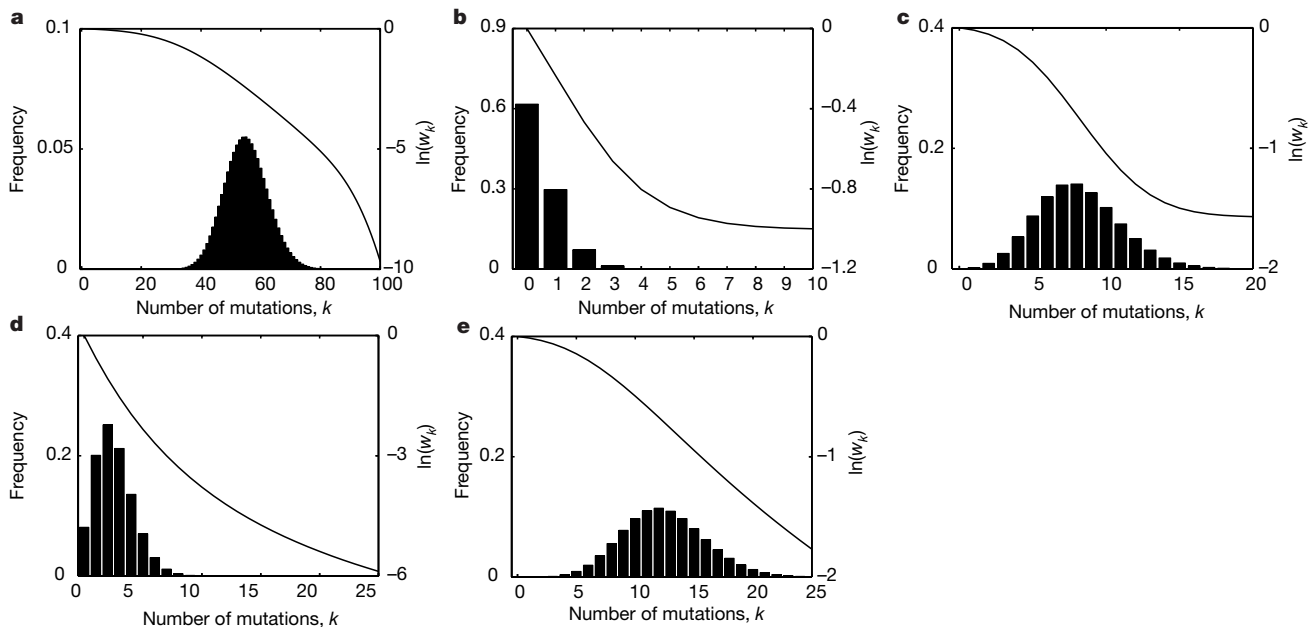


Figure 1 Results of five numerical studies. In each case, the histogram gives the equilibrium frequencies of adults with k mutations, where $k = 0, 1, 2, 3, \dots$. The curve gives the natural logarithm of the w_k values. The value of w_k is the expected number of offspring for an adult with k deleterious mutations divided by the expected number of offspring for an adult with no deleterious mutations (measured at equilibrium). The parameter values used are: **a**, $\alpha = 0.2$, $\Theta = 100$, $\gamma = 0.01$, $U = 4.5$, $C_k = 0.9^k$; **b**, $\alpha = 0.04$, $\Theta = 100$, $\gamma = 10^{-9}$, $U = 0.1$, $C_k = 0.5^k$; **c**, $\alpha = 0.04$, $\Theta = 100$, $\gamma = 10^{-9}$, $U = 1.0$, $C_k = 0.5^k$; **d**, $\alpha = 0.04$, $\Theta = 2,000$, $\gamma = 0.01$, $U = 1.0$,

$C_0 = 1$, $C_k = \prod_{i=1}^k [0.95 - 0.4e^{-0.1(i-1)}]$ for $k \geq 1$; **e**, $\alpha = 0.04$, $\Theta = 100$, $\gamma = 0.01$, $U = 1.0$, $C_0 = 1$, $C_k = \prod_{i=1}^k [0.95 - 0.4e^{-0.1(i-1)}]$ for $k \geq 1$. At equilibrium, the values of $\bar{n}$ (the mean number of adults consuming a food item), $\bar{k}$ (the mean number of mutations per adult) and $\bar{w}$ (the mean of w_k) are: **a**, $\bar{n} = 6.98$, $\bar{k} = 55.057$, $\bar{w} = 0.104$; **b**, $\bar{n} = 1.594$, $\bar{k} = 0.483$, $\bar{w} = 0.904$; **c**, $\bar{n} = 1.594$, $\bar{k} = 8.117$, $\bar{w} = 0.512$; **d**, $\bar{n} = 39.973$, $\bar{k} = 2.549$, $\bar{w} = 0.355$; **e**, $\bar{n} = 0.400$, $\bar{k} = 12.446$, $\bar{w} = 0.508$.

Members of our hypothetical population spend their time looking for and consuming food items. Once an adult encounters and begins to consume a food item, consumption continues until the food item is gone. The quantity of food in a food item decreases as a result of both consumption and decay. Decay occurs at rate γ (where $\gamma > 0$). For convenience, we assume that, when consumption begins, all food items contain α units of food. Thus, if n adults consume a food item at the same time, and they have a mean rate of consumption $\bar{C}$, then the food item will disappear after $\alpha/(n\bar{C} + \gamma)$ time units (where $n \geq 1$). In this case, the amount of a food item

that is consumed by an adult with k mutations is $\alpha C_k/(n\bar{C} + \gamma)$.

Let Θ denote the expected value for the number of food items that an adult will encounter between maturation and death. We assume that variation among adults in the value of Θ is negligible, and so treat Θ as a constant. This suggests that adults spend a negligible amount of their time consuming food, compared to the amount of time spent looking for food.

The number of adults that consume each food item is assumed to follow a Poisson distribution with mean $\bar{n}$. Thus, a proportion $\exp(-\bar{n})$ of food items are never consumed. The individuals that

Box 1

The main mathematical details of the analysis

The population is assumed to be sufficiently large that stochastic effects may be ignored. In one particular generation, p_k denotes the frequency of adults with k mutations. Quantities associated with the following generation are indicated by a prime. Results are presented here only for the sexual population (results for the asexual population follow straightforwardly).

We census adults immediately before the production of gametes. The expected number of offspring produced by a female with k mutations is F_k .

The rule connecting $\bar{n}$ in adjacent generations is $\bar{n}' = \bar{n}\bar{F}/2$, where $\bar{F} = \sum_{k=0}^{\infty} F_k p_k$. Thus at equilibrium, when $\bar{n}' = \bar{n}$, the expected number of offspring is two for every female.

No runaway theorem for $\bar{k}$

We shall establish here the result that for the sexual population, the mean number of mutations, $\bar{k} = \sum_{k=0}^{\infty} k p_k$, cannot increase indefinitely. Thus if, in one generation, $\bar{k}$ is sufficiently large, then for the following generation, $\bar{k}' < \bar{k}$.

It follows from the dynamical equation relating p_k in adjacent generations that

$$\bar{k}' = U + \sum_{k=0}^{\infty} k F_k p_k / \bar{F} \quad (2)$$

Let us first analyse this equation when $\gamma = 0$. Although this value of γ is not allowed within the model, it will provide useful information for the allowed case: $\gamma > 0$.

For $\gamma = 0$, it can be shown that $\bar{F} = \alpha\Theta[1 - \exp(-\bar{n})]/\bar{n}$, and equation (2) can be written as

$$\bar{k}' - \bar{k} = U + \frac{\alpha\Theta}{\bar{F}} \sum_{m=0}^{\infty} \frac{Q(m)}{m+1} \sum_{k_0, k_1, \dots, k_m} \left(\frac{\sum_{j=0}^m k_j C(k_j)}{\sum_{j=0}^m C(k_j)} - \frac{\sum_{j=0}^m k_j}{\sum_{j=0}^m 1} \right) p_{k_0} p_{k_1} \dots p_{k_m} \quad (3)$$

where $Q(m) = \bar{n}^m \exp(-\bar{n})/m!$. For all m , the expression on the right-hand side of equation (3) inside brackets is negative (using Chebyshev's inequality: see equation (11.115) of ref. 30). Furthermore, the leading non-zero term in the sum over m on the right-hand side of equation (3) occurs at $m = 1$. This is the contribution from two adults encountering and consuming a food item. Assuming that the mean number of adults per food item, $\bar{n}$, is finite, it follows that $Q(1)$ is finite and a non-negligible proportion of food items are encountered by two adults. We shall show that when $\bar{k}$ is sufficiently large in one generation, the contribution from just the $m = 1$ term on the right-hand side of equation (3) is sufficiently negative that $\bar{k}'$ will be smaller than $\bar{k}$. As a consequence, the mean number of mutations cannot increase indefinitely with time in a sexual population.

We separate the possible distributions of mutations among adults into two cases that cover all possibilities.

Case (1). The proportion of all adults, in one generation with $< \bar{k}/4$ mutations is $\leq \lambda$ (where $\lambda \neq 0, 1$).

In the next generation, in which $\bar{k}$ is calculated, a proportion $\geq (1 - \lambda)^4$ of the food items consumed by two adults will be eaten by individuals whose parents both had $\geq \bar{k}/4$ mutations.

The gametes produced by two parents with a large number of mutations will be normally distributed with a mean of one-half the parental number of mutations and a variance of one-quarter the parental number of mutations. Thus the mean number of mutations in offspring produced

from two parents with $\geq \bar{k}/4$ mutations is $\geq \bar{k}/4$ and the variance in the number of mutations in these offspring is $\geq \bar{k}/8$.

The contribution to $\bar{k}' - \bar{k}$ from the $m = 1$ term of equation (3) can be written

$$- \frac{\alpha\Theta Q(1)}{4\bar{F}} \sum_{k_0, k_1} |k_1 - k_0| \frac{|C(k_1) - C(k_0)|}{C(k_1) + C(k_0)} p_{k_0} p_{k_1} \quad (4)$$

The contribution to this from the offspring of two parents with $\geq \bar{k}/4$ mutations is of order $-\alpha\Theta Q(1)(1 - \lambda)^4 \sqrt{\bar{k}/\bar{F}}$ (more refined estimates can be made, but the result given is adequate for our purposes). Thus for sufficiently large $\bar{k}$, this contribution becomes large and negative and

forces $\bar{k}' - \bar{k}$ to be negative, thereby preventing runaway of the mean number of mutations.

Case (2). The proportion of all adults, in one generation, with $< \bar{k}/4$ mutations is $> \lambda$ (where $\lambda \neq 0, 1$).

In this case, some adults have $< \bar{k}$ mutations and there must be other adults with $> \bar{k}$ mutations. Let the proportion of the adults with $> \bar{k}$ mutations be R . We focus on food items that are consumed by exactly two adults where both parents of one adult had $< \bar{k}/4$ mutations and the other adult had at least one parent with $> \bar{k}$ mutations. Under the current assumptions, the frequency with which consumed food items fall into this category is at least $2R(1 - R)\lambda^2$. Furthermore, when $\bar{k}$ is very large, the difference in the number of mutations of adults on all but a negligible proportion of such food items must be at least of the order of $\bar{k}$. As a consequence, the contribution to $\bar{k}' - \bar{k}$ arising from such food items is, from equation (4), of order $-\alpha\Theta Q(1)R(1 - R)\lambda^2 \sqrt{\bar{k}/\bar{F}}$. For sufficiently large $\bar{k}$, this contribution becomes large and negative and forces $\bar{k}' - \bar{k}$ to be negative.

Cases (1) and (2) indicate that for $\gamma = 0$, $\bar{k}$ does not $\rightarrow \infty$. Provided all consumers of a food item have a finite number of mutations, the behaviour in γ is continuous as $\gamma \rightarrow 0$. Thus for $\gamma \rightarrow 0$ the $\gamma = 0$ results are obtained.

Other results

A number of other results are contained in the main text. Here we provide details of their derivation.

The equilibrium value of $\bar{n}$ for $\gamma \rightarrow 0$, is given by the positive solution of $2\bar{n} = \alpha\Theta[1 - \exp(-\bar{n})]$. This follows when $\bar{F} = \alpha\Theta[1 - \exp(-\bar{n})]/\bar{n}$ (which applies for $\gamma \rightarrow 0$), is used in the update rule for $\bar{n}$ at equilibrium: $\bar{n}' = \bar{n}\bar{F}/2$.

The condition for invasion of a sexual population by asexuals follows by first noting the number of offspring produced by a mutation-free asexual invader. This number is identical to that of a mutation free sexual, namely F_0 , because the asexual is in competition only with sexuals for food. The critical mutation rate U_1^* is that for which a mutation-free asexual is just able to replace itself with another mutation free individual, given by $\exp(-U_1^*)F_0 = 1$. As $\bar{W} = \bar{F}/F_0$ and at equilibrium $\bar{F} = 2$, it follows that U_1^* is determined by $\exp(-U_1^*)(2\bar{W}) = 1$, which is equivalent to the expression given in the main text.

In the modified model, where there is a maximum value of $\bar{n}$, namely $\bar{n}_{\max}$, we note that $F_0 < \alpha\Theta$ and individuals are expected to encounter at least $\Theta \exp(-\bar{n}_{\max})$ food items that are not encountered by any other individual. Thus $\bar{F} > \alpha\Theta \exp(-\bar{n}_{\max})$, so $\bar{W} > \exp(-\bar{n}_{\max})$, and we have the bound $U_1^* < \bar{n}_{\max} + \ln(2)$.

consume a particular food item are a random sample of the adults, and we take $\bar{n}$ to be proportional to the number of adults in the population (which is assumed to be very large). In our numerical studies we initially take $\bar{n} = \alpha\Theta$, and initially all adults have no mutations.

Extensive numerical study suggests that this model always leads to equilibrium. We determine the w_k values only after equilibrium is reached.

Consider the case where, at equilibrium, the mean number of individuals per food item ($\bar{n}$) is very large ($\bar{n} \rightarrow \infty$). In this case, we can show that w_k is proportional to C_k for all $k \geq 0$. Thus, when $\bar{n}$ is very large, there is no a priori reason to expect synergistic epistasis to emerge from competition for resources.

A very large value of $\bar{n}$ can only be achieved if the availability of food ($\alpha\Theta$) is also very large. When food availability is more limited a variety of patterns are possible (Fig. 1). Figure 1a–c uses multiplicative C_k functions. This sort of function leads to the logarithm of w_k declining linearly with k when $\bar{n}$ is very large. However, when food availability is limited, multiplicative C_k functions can lead to synergistic epistasis (Fig. 1a), diminishing-returns epistasis (Fig. 1b) or a combination of the two (Fig. 1c).

Despite the variety of patterns that arise when food availability is limited, it is possible to say something general in the case where γ (the rate of food decay) is very small ($\gamma \rightarrow 0$). In this case, we can show that for k sufficiently large (but finite) we have $w_{k+1}/w_k \approx 1$. In other words, if we consider individuals with sufficiently large numbers of mutations, then any additional mutations have a negligible effect on fitness. We can expect the effect of additional mutations on fitness to be non-negligible for some smaller values of k . Thus, when γ is very small, we will have diminishing-returns epistasis for sufficiently large values of k (Fig. 1b, c). These findings arise because adults with many more mutations than average will consume almost no food from food items that are consumed by multiple adults (because these food items will be consumed very quickly, relative to the rate of consumption for such adults). However, if γ is very small, these mutation-laden adults will consume almost all of any food items that they alone encounter.

Another case where it is possible to say something general about the shape of the w_k curve is where γ is very small ($\gamma \rightarrow 0$), $\bar{n}$ is not very large and nearly all members of the population have such large numbers of mutations that they consume food items very slowly in comparison with individuals with few mutations. In this case, adults with only a few deleterious mutations will generally consume food items very quickly in comparison with the rate of food decay, and in comparison with any other adults that are consuming the same food item. Therefore, these adults will usually consume nearly all of any food item they encounter. Thus, $w_{k+1}/w_k \approx 1$ for sufficiently small values of k . For values of k closer to the mean we can expect smaller values of w_{k+1}/w_k because of competition with other adults. Thus, regardless of the details of the pattern of C_k values, synergistic epistasis can be expected for sufficiently small values of k (Fig. 1a, c, e).

This mechanism for producing synergy requires a sufficiently small value of $\bar{n}$, and so food availability ($\alpha\Theta$) must be sufficiently small. With this in mind, consider Fig. 1d and e which uses a function for C_k that gives diminishing-returns epistasis when food availability is sufficiently large. For Fig. 1d we set $\Theta = 2,000$, leading to an equilibrium where $\bar{n} = 39.973$. As a consequence of this substantial $\bar{n}$ value, the w_k values shown are similar to the C_k values (all are within 8%), and diminishing-returns epistasis prevails. When food availability is cut to $\Theta = 100$, $\bar{n} = 0.400$ at equilibrium. As a consequence, the w_k values show synergistic epistasis (Fig. 1e).

Another possible source of synergistic epistasis relates to food decay. In a sense, adults must compete against food decay in a way that is similar to the competition they experience with each other. However, this source of synergy is unimportant when γ is low

relative to the average rate of food consumption. This is easily demonstrated. For example, decreasing the value of γ from 0.01 to 10^{-9} does not cause much change in the w_k values in Fig. 1a (data not shown).

The population described above will survive only if females consume enough food to produce at least two offspring each (on average). In the limit as $\gamma \rightarrow 0$ the condition for survival of a sexual population is $\alpha\Theta > 2$. This means that, when $\alpha\Theta > 2$, a sexual population can survive any finite rate of mutation (U) so long as γ is sufficiently small. (Proofs are given in Box 1). We also find that when $\alpha\Theta > 2$ and we take the limit as $\gamma \rightarrow 0$, the equilibrium population density takes on a value such that the equilibrium value of $\bar{n}$ is given by the positive solution of $2\bar{n} = \alpha\Theta[1 - \exp(-\bar{n})]$. Note that, in this case, U has no influence on the equilibrium population density. These results arise because, when γ is very small, population extinction cannot occur unless the mean number of mutations per adult ($\bar{k}$) is very large, so that the average rate of food consumption is depressed at least to the point where food consumption occurs at a similar rate to food decay (γ). A low γ and a high $\bar{k}$ are exactly the conditions specified above for the emergence of synergistic epistasis over sufficiently small values of k . Furthermore, these conditions tend to produce situations where $w_k \approx 1$ for all but very large values of k . Functions of this sort are known to be particularly beneficial for sexual populations¹.

Next, consider the limit as $U \rightarrow 0$. In this limit, the condition for survival of a sexual population is $\alpha\Theta > 2(1 + \gamma/C_0)$.

What happens to a sexual population when $U \gg 0$ and $\gamma \gg 0$? The answer depends on the exact choice of parameter values. However, so long as $\alpha\Theta > 2$ it is not difficult to find parameter values for which a sexual population can survive, but an asexual population (described below) will go extinct. This can be done by choosing γ sufficiently small and U sufficiently large (see Fig. 2 for examples).

Let us consider an asexual population that is identical to the one just described except that all individuals are female, and each offspring is genetically identical to its parent, except for new mutations. Mutations occur at the same rate per locus as in the

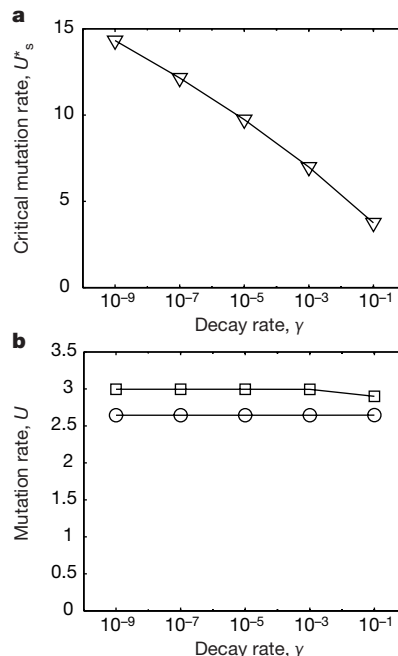


Figure 2 Results from the numerical studies. **a**, The values of U_s^* , the genomic mutation rate above which a sexual population goes extinct. **b**, Squares, values of U_s^* , the genomic mutation rate above which an asexual population will go extinct. Circles, values of U_i^* , the genomic mutation rate above which a sexual population at equilibrium cannot be successfully invaded by an asexual lineage. For all data shown, $\alpha = 0.2$, $\Theta = 100$, $C_k = 0.9^k$.

sexual population described above, and thus the mean number of new mutations per genome is U .

This asexual population will perpetuate itself indefinitely so long as the average amount of food consumed per adult is sufficient to produce at least one offspring. This is half the food requirements of a sexual population, which reflects the well known two-fold cost of sex²⁵. Over the long term, the rate of food consumption can be maintained at a sufficiently high level only if

$$\frac{\alpha\Theta e^{-U}}{1 + \gamma/C_0} \geq 1 \quad (1)$$

If this relation is not satisfied, the population will go extinct. This result is in sharp contrast to the situation in sexuals. Whenever $e^U > \alpha\Theta > 2$ and γ is sufficiently small, we have a situation where a sexual population can survive but an asexual population will go extinct (for any allowable values of C_k). For example, if γ is very small, then this is the case when $\alpha\Theta = 2.5$ and $U > 0.917$, and also when $\alpha\Theta = 25$ and $U > 3.22$.

Let us consider the conditions required for a sexual population to be safe from invasions into its habitat by asexual populations that consume the same food. The results will also tell us when a sexual population is safe from mutations that cause asexual reproduction and eliminate the 'cost of sex' by causing female adults to produce only female offspring. Mutations (or sets of mutations) of this sort seem to be responsible for the origination of many asexual populations^{26–28}.

Suppose that the invading asexuals are as described above, and have the same mutation rate and C_k values as the sexuals. We assume that the asexuals are initially very rare, so that they compete with sexuals for food, but they rarely compete with each other for the same food item. Under these conditions, we find that a sexual population at equilibrium cannot be invaded by asexuals so long as the genomic mutation rate (U) is sufficiently large. Let U_1^* represent the value of U above which asexuals can never invade by increasing from a very low initial density (I is for invasion). If $U < U_1^*$ then invading asexuals may increase to a non-negligible density (such an increase is guaranteed if $U < U_1^*$ and many of the asexual invaders are free of deleterious mutations). It can be shown that $U_1^* = \ln(2/\bar{w})$, where $\bar{w}$ is the mean value of w_k . The values of w_k and $\bar{w}$ can be derived from numerical study of a sexual population (see Fig. 2 for an example). We can also show that, in all cases, $U_1^* \leq \ln(\alpha\Theta)$.

In our model, food availability has a strong effect on population density. In many natural populations, however, population density is largely determined by other factors, such as predation and parasitism²⁹. We can modify the model to emulate these processes in a very rough way by assuming that the mean number of adults per

food item ($\bar{n}$) can never exceed a particular value, denoted $\bar{n}_{\max}$. Assume that, if the population at birth is such that the number of newborn individuals per food item is greater than $\bar{n}_{\max}$, then a 'thinning' process takes place so that randomly selected newborns die until $\bar{n} = \bar{n}_{\max}$.

If $\bar{n}_{\max}$ is sufficiently large then thinning never occurs. What happens, however, if $\bar{n}_{\max}$ is small enough so that thinning does occur? The condition for survival of an asexual population is unaffected, and is still given by relation (1). Furthermore, a sexual population can still survive so long as $\alpha\Theta > 2$ and γ is sufficiently small. Perhaps the most intriguing effect is the impact on U_1^* . For any choice of parameter values such that $\alpha\Theta > 2$, we have, in the limit as $\gamma \rightarrow 0$, $U_1^* < [\bar{n}_{\max} + \ln(2)]$. Thus, if γ is sufficiently small, we can bring U_1^* arbitrarily close to $\ln(2) = 0.693$ by decreasing the value of $\bar{n}_{\max}$. This is because when γ is sufficiently small and $\bar{n}_{\max}$ is also small, competitions over food items are rare, and so the fitness of mutation-free individuals is not much higher than the mean fitness. This limits the maximum possible advantage of mutation-free asexual invaders.

Figure 3 shows data for a set of numerical studies that used the same parameter values used for Fig. 2, except that $\bar{n} = 1.386$, so that 25% of the food items that adults consume are consumed by no other adult. Comparison of Fig. 2 with Fig. 3 shows that the imposition of a maximum population density causes a substantial decrease in U_1^* in this case. With the preceding analytic result, this suggests that, when γ is sufficiently small, sexual populations will tend to be best protected from invasions by asexuals when their population density is kept low enough to allow many food items to be consumed by only one adult.

Our results show that synergistic epistasis can emerge from competition for food. Similar models can be constructed to show that synergy can arise from competition for other limited resources such as nest sites, places in which to shelter, distance from predators, access to light (in plants) and so on. In sexual organisms, the interactions involved in competing for mates may be a particularly important cause of synergistic epistasis. Additionally, deleterious mutations are just one source of variation in fitness. It is reasonable to expect results similar to those produced here for other sources of variation (for example, environmental change or spatial heterogeneity and migration).

To our knowledge, there has never been an experimental search for synergy that has allowed members of a genetically diverse population to compete in many randomly formed groups. Our results indicate that without such an approach, there may never be a convincing demonstration of the reality of synergistic epistasis. Our results also suggest that experimenters should be cautious, because, as shown in Fig. 1, synergistic epistasis may not hold over the entire range of mutation-contamination levels observed within a population. □

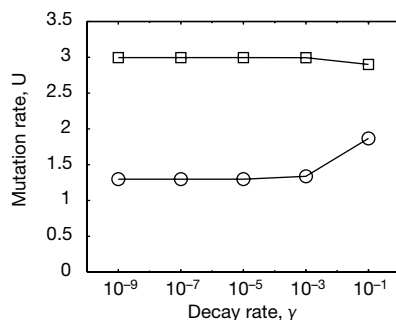


Figure 3 Results from the numerical studies using the modified model, for which there is a maximum allowable value of $\bar{n}$. Squares, values of U_1^* , the genomic mutation rate above which an asexual population will go extinct. Circles, values of U_2^* , the genomic mutation rate above which a sexual population at equilibrium cannot be successfully invaded by an asexual lineage. The U_2^* values are identical to those shown in Fig. 2. For all data shown, $\bar{n}_{\max} = 1.386$, $\alpha = 0.2$, $\Theta = 100$, $C_k = 0.9^k$.

Received 11 January; accepted 9 May 2000.

- Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440 (1988).
- Charlesworth, B. Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet. Res.* **55**, 199–221 (1990).
- Peck, J. R., Barreau, G. & Health, S. C. Imperfect genes, Fisherian mutation and the evolution of sex. *Genetics* **145**, 1171–1199 (1997).
- Charlesworth, B. Directional selection and the evolution of sex and recombination. *Genet. Res.* **61**, 205–224 (1993).
- Waxman, D. & Peck, J. R. Sex and adaptation in a changing environment. *Genetics* **153**, 1041–1053 (1999).
- Kondrashov, A. S. Muller's ratchet under epistatic selection. *Genetics* **136**, 1469–1473 (1994).
- Kondrashov, A. S. Contamination of the genome by very slightly deleterious mutations—why have we not died 100 times over? *J. Theor. Biol.* **175**, 583–594 (1995).
- Butcher, D. Muller's ratchet, epistasis and mutation effects. *Genetics* **141**, 431–437 (1995).
- Kimura, M. & Maruyama, T. The mutational load with epistatic gene interactions in fitness. *Genetics* **54**, 1337–1351 (1966).
- Crow, J. F. in *Mathematical Topics in Population Genetics* (ed. Kojima, K.) (Springer, Berlin, 1970).
- de Visser, J. A. G. M., Hoekstra, R. F. & van den Ende, H. An experimental test for synergistic epistasis and its application in *Chlamydomonas*. *Genetics* **145**, 815–819 (1997).
- de Visser, J. A. G. M. & Hoekstra, R. F. Synergistic epistasis between loci affecting fitness: evidence in plants and fungi. *Genet. Res.* **71**, 39–49 (1998).

13. Elena, S. F. & Lenski, R. E. Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**, 395–398 (1997).
14. Koelwijn, H. P. Effects of different levels of inbreeding on progeny fitness in *Plantago coronopus*. *Evolution* **52**, 692–702 (1998).
15. West, S. A., Peters, A. D. & Barton, N. H. Testing for epistasis between deleterious mutations. *Genetics* **149**, 435–444 (1998).
16. Willis, J. H. Effects of different levels of inbreeding on fitness components in *Mimulus guttatus*. *Evolution* **47**, 864–876 (1993).
17. Eyre-Walker, A. & Keightley, P. D. High genomic deleterious mutation rates in hominids. *Nature* **397**, 344–347 (1999).
18. Eyre-Walker, A. Evidence of selection on silent site base composition in mammals: Potential implications for the evolution of isochores and junk DNA. *Genetics* **152**, 675–683 (1999).
19. Fry, J. D., Keightley, P. D., Heinsohn, S. L. & Nuzhdin, S. V. New estimates of the rates and effects of mildly deleterious mutation in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **96**, 574–579 (1999).
20. Kondrashov, A. S. & Turelli, M. Deleterious mutations, apparent stabilizing selection and the maintenance of quantitative variation. *Genetics* **132**, 603–618 (1992).
21. Peck, J. R. & Eyre-Walker, A. The muddle about mutations. *Nature* **387**, 135–136 (1997).
22. Sved, J. A., Reed, T. E. & Bodmer, W. F. The number of balanced polymorphisms that can be maintained in a natural population. *Genetics* **55**, 469–481 (1967).
23. King, J. L. Continuously distributed factors affecting fitness. *Genetics* **55**, 483–492 (1967).
24. Hamilton, W. D., Axelrod, R. & Tanese, R. Sexual reproduction as an adaptation to resist parasites (a review). *Proc. Natl Acad. Sci. USA* **87**, 3566–3573 (1990).
25. Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, Cambridge, 1978).
26. Parker, E. D. in *Evolutionary Genetics from Molecules to Morphology* (eds Singh, R. & Krimbas, C.) 450–475 (Cambridge Univ. Press, Cambridge, 2000).
27. Nygren, A. Apomixis in the angiosperms. *Bot. Rev.* **20**, 577–649 (1954).
28. Bell, G. *The Masterpiece of Nature* (Univ. California Press, San Francisco, 1982).
29. Krebs, C. J. *Ecology* 4th edn (Harper Collins, New York, 1994).
30. Gradshteyn, I. S. & Ryzhik, I. M. *Table of Integrals, Series and Products* (Academic, New York, 1980).

Acknowledgements

We thank J. Crow, A. Eyre-Walker, A. Kondrashov and J. Maynard Smith for advice. This study was supported by the Biotechnology and Biological Sciences Research Council (UK).

Correspondence and requests for materials should be addressed to J.R.P. (e-mail: j.r.peck@sussex.ac.uk).

Long-term vocal recognition in the northern fur seal

Stephen J. Insley

Animal Behavior Group, University of California, Davis, California 95616 and Department of Zoological Research, National Zoological Park, Smithsonian Institution, Washington DC 20008, USA

The ability to recognize and remember individual identities for long periods of time has important implications for the evolution of animal social behaviour, particularly complex interactions such as cooperation or mate choice^{1–7}. Despite this importance, there is only a single example of long-term individual recognition in nature, the 8-month retention of neighbour's song among male hooded warblers, *Wilsonia citrina*⁷, and there is none for a non-human mammal. Associations between individuals spanning years, which are especially prevalent in carnivores⁸, primates⁹ and seabirds¹⁰, and evidence of mate fidelity^{11,12} provide indirect support for the ability of long-term recognition. In many of these instances, however, individuals do not separate for extended periods, and thus long-term recognition, although often assumed, may be both unnecessary and nonexistent. Furthermore, site fidelity rather than individual recognition may explain many instances of mate fidelity¹⁰. Here I show that mother–offspring pairs of a migratory otariid pinniped—the northern fur seal (*Callorhinus ursinus*)—not only have the ability to recognize each other's vocalizations during the course of a breeding season, but are also able to retain these memories for at least 4 years.

Vocal recognition is a key component of offspring survival among northern fur seal maternal dyads and consequently both mothers and pups are very responsive to vocal playback studies^{13,14}.

Throughout nursing, mothers leave on week-long foraging trips and upon their return must relocate their pups in dense breeding colonies using what appears to be a combination of geographic, vocal and olfactory cues. At four months of age, pups migrate south independent of their mothers^{15,16}. The seals spend winter at sea, travelling as far south as the Channel Islands, California, and as far east as Japanese waters¹⁵. Given the large distances travelled, and the different migratory habits of immature and adult seals^{15,16}, it is unlikely that parents and immature offspring would encounter one another while away from the breeding areas. When seals return they tend to frequent their natal sites¹⁷, making it possible for individuals to meet again in future seasons. Females are sexually mature between 3 and 6 years of age, having their first potential offspring during year 4 after a 1-year delayed implantation and gestation. Male sexual maturation also occurs between 3 and 6 years, but males are unable to compete for females until year 10 (ref. 15).

I tested the effect of time on mother–offspring vocal recognition using playback experiments that spanned three periods: within season (3–4-week delay), between seasons (1-year delay) and between several seasons (4-year delay). I tested within season effects because it is the most critical period for recognition, and concurrently, when pups experience maximal growth. If growth has a detrimental effect on the acoustic cues used for recognition, I would expect to see a decreased response to 'old' calls (where there was a 3–4-week lag between recording and playback) as compared with 'recent' calls (those with a 2–3-day lag). Responses to both of these treatments were expected to be greater than to 'control' calls (calls of familiar non-mothers and non-offspring). I found that both mother and pup vocal responses to the three treatments differed significantly (repeated measures analysis of variance (ANOVA) of mother's responses: $F_{2,10} = 3.78$, $P < 0.05$; and pup's responses, $F_{2,14} = 13.43$, $P < 0.0001$). Whereas recent and old pup calls elicited more responses from mothers than control pup calls, responses to these two categories were not distinguishable from each other (paired *t*-tests: recent compared with control, $t_{10} = 2.97$, $P < 0.05$; old compared with control, $t_{10} = 2.96$, $P < 0.05$; recent compared with old, $t_{10} = 0.28$, $P > 0.05$). The same pattern was true for pup responses to their mother's calls (paired *t*-tests: recent compared with control, $t_{14} = 4.18$, $P < 0.001$; old compared with control, $t_{14} = 4.74$, $P < 0.001$; recent compared with old, $t_{14} = 1.83$, $P > 0.05$). This experiment shows that a short time lag of 3–4 weeks, double the maximum separation time naturally experienced during female foraging trips when lactating, does not affect vocal recognition.

To test whether recognition was maintained over consecutive seasons, I measured mother and pup vocal responses to playbacks of their offspring or mother's calls recorded the previous year. Test treatments were matched with year-old control calls. I found that both mothers and pups responded more to the year-old calls of their offspring or their mothers than to control calls (paired *t*-tests: mother responses, $t_5 = 2.7$, $P < 0.05$; pup responses, $t_5 = 6.71$, $P < 0.01$) showing that vocal recognition can be maintained for at least 1 year.

To test recognition after multiple seasons, I conducted playbacks to 4-year-old females during their first observed visits to the breeding area since they were neonates. I was able to measure the responses of four such females to playbacks of their mother's calls made when the subjects were pups. As before, test treatments were matched with similar aged control calls. None of the females had an offspring, appeared pregnant, or was seen to associate with an older female that could have been her mother. Each one remained on land for a maximum of 3 days and afterwards was not observed again. Each of the four females orientated to the test treatments and not to the control treatments (paired sign test: $P < 0.01$, $n = 8$; 2 playbacks to each of 4 subjects¹⁸). Two of the four females also responded vocally to the playbacks (insufficient for statistical comparisons). These two females called only to the test treatments and not once to the control treatments (the first female called during only one trial, and the second called during three trials). The specific vocal

responses observed here rarely occur in young females without offspring, and, apart from during the playback experiments, none of the subject females were observed calling.

Long-term individual recognition could provide fitness benefits through cooperative or choice facilitating associations between parents and mature offspring, cohort members, mating pairs and neighbouring territory holders in a variety of species^{1,3,6,7}. The sensory mechanism used for long-term recognition must sufficiently counteract morphological development and accurately store and recall information from long-term memory. My experiments show that vocal recognition can be stable over time, even through periods of maximal growth, and that memory and recall of vocalizations can be maintained for several years with little or no updating. Although recognition performance is probably enhanced by signal redundancy in other sensory modalities (for example, olfaction and vision), the acoustic/auditory channel alone is clearly sufficient for long-term individual recognition in northern fur seals. □

Methods

Behavioural studies of northern fur seals have been carried out at the same site on St. Paul Island, one of the Pribilof Islands in Alaska, since 1990. Most seals in the study cove (~5 adult males and 30–40 mother–offspring pairs) were marked each year with hair bleach for within season identification, and a limited number of flipper tags were applied (50 pups and 44 adult females) for long-term identification. The calls used for each of the three playbacks are broadcast vocalizations, commonly referred to as 'attraction calls', the most frequently used call by adult females and neonates. Each experimental treatment was made up of a group of five different call exemplars spaced at 1-s intervals (a natural repetition rate¹³) and played back with 30 s between groups. Initial treatment orders were randomized and thereafter alternated. All control treatments used familiar calls (calls from seals in the same study cove whose calls were clearly audible to the experimental subjects), maximizing the possibility that individual recognition was being tested for and not simply differential responsiveness to new and familiar calls. Detailed procedural accounts of marking, recording/processing of vocalizations and playback methods are provided elsewhere¹⁴.

Each of the three playbacks presented the treatments sequentially; experiment one used a repeated measures design with three treatments, and experiments two and three used a switchback design with two treatments. For the first experiment, the recent call exemplars were recorded less than 1 week from the time of the playback and were mostly 2–3 days old (the duration of a mother's time on land between foraging trips). Old call exemplars were recorded 3–4 weeks before playback presentations (equivalent to 2–4 female foraging trips). Control calls were matched for age. In addition, both mothers and pups had always experienced the current state of the other's calls before playbacks were carried out; that is, playbacks were not made to mothers upon their arrival back from a foraging trip, before they had contacted their pups, using recent call versions they had not yet experienced. For the second experiment, the test treatments played to mothers were calls of their pup from the previous year that had been recorded while these pups were still suckling. The test treatments played to pups were their mother's calls recorded in the previous year, before the subject pups were born. Thus, the year-old playbacks to pups test the acoustic stability of adult female calls over a period of 1 year. For the third experiment, the test treatments played to 4-year-old females were their mother's calls recorded 4 years earlier while the subjects were dependent pups.

Each playback experiment was scored with three behavioural assays on site: vocalizations, orientation and phonotaxis (movement towards the playback speakers). Scores of each playback were independently verified afterwards with videotaped records. Only experiments in which the subject gave at least one response (vocal or behavioural to either control or test treatments) were used in the analyses. Vocal responses provide the clearest assays and were sufficient to score the first two experiments. Vocal responses during experiment 3 were insufficient for statistical comparisons, and so the simple dichotomous response assay (whether or not subjects orientated towards the speakers) was used. Data were analysed with parametric statistics except for experiment 3, which was not distributed normally. Degrees of freedom report the number of individual seals ($n - 1$) in all cases except for experiment 3. This experiment uses two contributions per subject as a means of dealing with low statistical power¹⁸. All statistical comparisons are two-tailed.

Received 14 April; accepted 2 June 2000.

1. Sherman, P. W., Reeve, H. K. & Pfennig, D. W. in *Behavioural Ecology: An Evolutionary Approach*, Fourth Edition (eds Krebs, J. R. & Davis, N. B.) 69–96 (Blackwell, Oxon, 1997).
2. Falls, B. J. in *Acoustic Communication in Birds, Volume 2: Song Learning and its Consequences* (eds Kroodsma, D. E. & Miller, E. H.) 237–278 (Academic, New York, 1982).
3. Stoddard, P. K. in *Ecology and Evolution of Acoustic Communication in Birds* (eds Kroodsma, D. E. & Miller, E. H.) 356–374 (Cornell Univ. Press, Ithaca, 1996).
4. Beecher, M. D. in *Kin Recognition* (ed. Hepper, P. G.) 94–124 (Cambridge Univ. Press, Cambridge, 1991).
5. Colgan, P. *Comparative Social Recognition* (Wiley, New York, 1983).
6. Mateo, J. M. & Johnston, R. E. Retention of social recognition after hibernation in Belding's ground squirrels. *Animal Behaviour* **59**, 491–499 (2000).
7. Godard, R. Long-term memory of individual neighbours in a migratory songbird. *Nature* **350**, 228–229 (1991).

8. Waser, P. M. in *Carnivore Behavior, Ecology, and Evolution, Volume 2* (ed. Gittleman, J. L.) 267–295 (Cornell Univ. Press, Ithaca, 1996).
9. Smuts, B., Cheney, D., Seyfarth, R., Wrangham, R. & Struhsaker, T. (eds) *Primate Societies* (Univ. Chicago Press, Chicago, 1987).
10. Black, J. M. *Partnership in Birds: The Study of Monogamy* (Oxford Univ. Press, New York, 1996).
11. Amos, B., Twiss, S., Pomeroy, P. & Anderson, S. Evidence for mate fidelity in the gray seal. *Science* **268**, 1897–1899 (1995).
12. Uy, J. A. C., Patricelli, G. L. & Borgia, G. Dynamic mate-searching tactic allows female satin bowerbirds *Ptilonorhynchus violaceus* to reduce searching. *Proc. R. Soc. Lond.* **267**, 251–256 (2000).
13. Insley, S. J. Mother-offspring separation and acoustic stereotypy: A comparison of call morphology in two species of pinnipeds. *Behaviour* **120**, 103–122 (1992).
14. Insley, S. J. Mother-offspring vocal recognition in northern fur seals is mutual but asymmetrical. *Animal Behaviour* (in the press).
15. Gentry, R. L. *Behavior and Ecology of the Northern Fur Seal* (Princeton Univ. Press, Princeton, 1998).
16. Regan, T., Antonelis, G. & Kyota, M. Early migration of northern fur seal pups from St. Paul Island, Alaska. *J. Mammal.* **76**, 1137–1148 (1995).
17. Baker, J. D., Antonelis, G. A., Fowler, C. W. & York, A. E. Natal site fidelity in northern fur seals, *Callorhinus ursinus*. *Animal Behaviour* **50**, 237–247 (1995).
18. Leger, D. W. & Didrichsons, I. A. An assessment of data pooling and some alternatives. *Animal Behaviour* **48**, 823–832 (1994).

Acknowledgements

This research was supported by the National Science and Engineering Research Council of Canada, the National Marine Fisheries Service and the American Museum of Natural History. We thank G. Antonelis, T. Steeves and B. Robson for help in the field; and P. Marler, A. Harcourt, D. Owings, J. Dumbacher, L. Shapiro, K. Phillips, T. Porter and D. Boness for helpful discussions on the manuscript.

Correspondence and requests for materials should be addressed to S.J.I. (e-mail: sinsley@nsp.si.edu).

State-dependent cross-inhibition between transmitter-gated cation channels

Baljit S. Khakh*, Xiaoping Zhou†, Jason Sydes*, James J. Galligan† & Henry A. Lester*

* Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

† Department of Pharmacology and Toxicology and the Neuroscience Program, Michigan State University, East Lansing, Michigan 48824, USA

Transmitter-gated cation channels are detectors of excitatory chemical signals at synapses in the nervous system¹. Here we show that structurally distinct $\alpha 3\beta 4$ nicotinic and P2X₂ channels influence each other when co-activated. The activation of one channel type affects distinct kinetic and conductance states of the other, and co-activation results in non-additive responses owing to inhibition of both channel types. State-dependent inhibition of nicotinic channels is revealed most clearly with mutant P2X₂ channels, and inhibition is decreased at lower densities of channel expression. In synaptically coupled myenteric neurons, nicotinic fast excitatory postsynaptic currents are occluded during activation of endogenously co-expressed P2X channels. Our data provide a molecular basis and a synaptic context for cross-inhibition between transmitter-gated channels.

Nicotinic receptors for acetylcholine ACh and the P2X receptors for ATP are transmitter-gated cation channels¹ that underlie cholinergic and purinergic fast synaptic transmission, and modulation of fast synaptic transmission. P2X and nicotinic receptors are structurally distinct¹ (Fig. 1a), and it has been assumed that they act independently^{2–4}. Studies of some neuronal nicotinic and P2X channels indicate, however, that these channels interact functionally^{5–8}. Overall, it is unclear whether P2X and nicotinic channels act independently in all cell types or in synapses expressing these receptors^{4–10}.

Native P2X receptors are heterogeneous⁹, but here we expressed P2X₂ and α3β4 channels in *Xenopus* oocytes because their properties resemble those of P2X and nicotinic channels expressed by peripheral neurons^{9–13}. ATP did not gate or modulate α3β4 channels; and conversely ACh did not gate or modulate P2X₂ channels (see Supplementary Information, Figs 1 and 2 and Table 1). In cells co-expressing P2X₂ and α3β4 channels (P2X₂ + α3β4), application of either ATP or ACh evoked characteristic inward currents; however, $I_{ATP/ACh}$ was smaller than the predicted sum of I_{ATP} and I_{ACh} (Fig. 1; $n = 33$, the predicted waveform is the arithmetic sum of the

I_{ATP} and I_{ACh} waveforms). P2X₂ and α3β4 nicotinic receptor currents were equally non-additive when the Na⁺ electrochemical driving force was reduced by lowering Na⁺ concentration or by measuring currents closer to the reversal potential (Table 1), indicating that non-additivity did not arise from poor voltage clamp. In cells co-expressing P2X₂ and α3β4 channels, the increase in membrane slope conductance (Δg , –10 mV) for I_{ATP} and I_{ACh} was $92 \pm 9 \mu S$ and $60 \pm 9 \mu S$, respectively, whereas that for $I_{ATP/ACh}$ was $98 \pm 8 \mu S$; but the predicted Δg for $I_{ATP/ACh}$ was $151 \pm 17 \mu S$ ($n = 3$). The P2X₂ channel ATP concentration-response curve

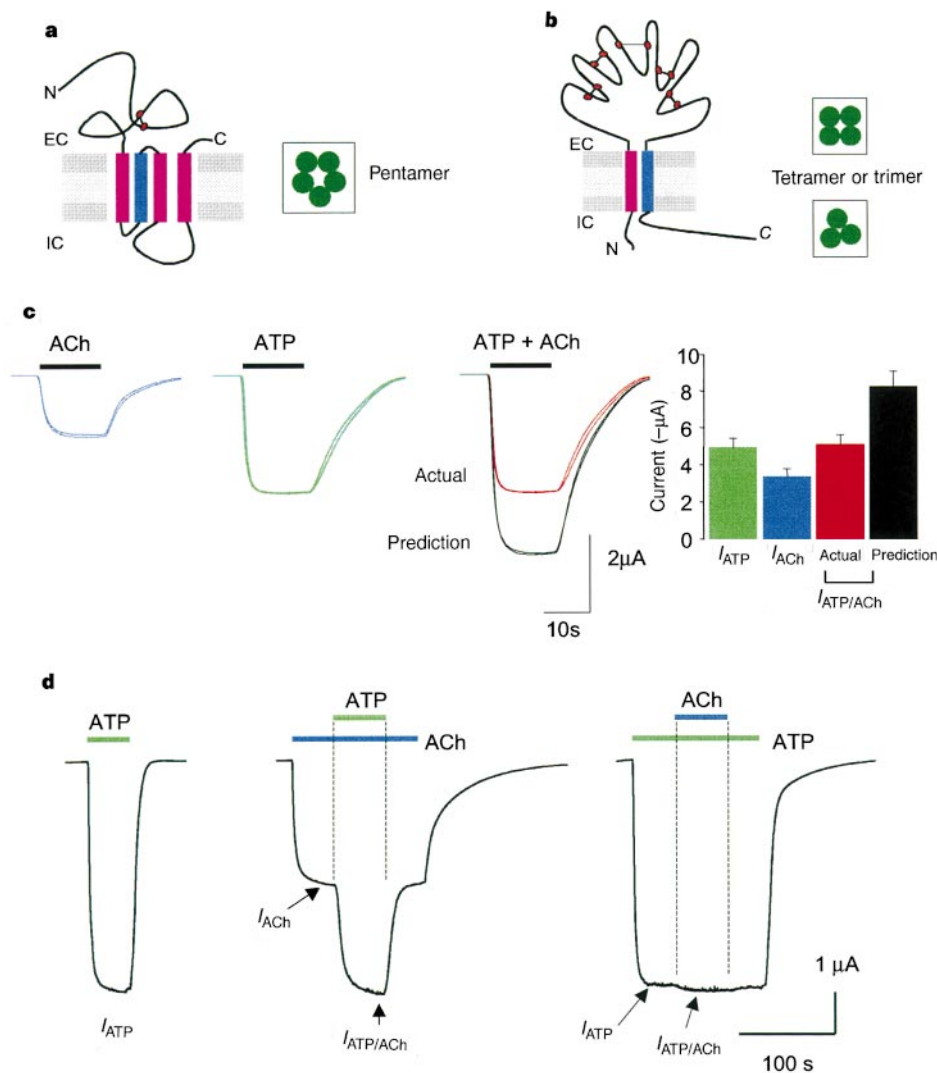


Figure 1 Non-additivity between P2X₂ and α3β4 nicotinic channels. **a**, **b**, Nicotinic (**a**) and P2X (**b**) receptor subunit topology and stoichiometry^{1,27}; blue indicates pore-lining segment. **c**, **d**, Inward currents from *Xenopus* oocytes expressing P2X₂ and α3β4 channels. **c**, Currents evoked with 100 μM ATP, 100 μM ACh, or a mixture of 100 μM ATP/ACh. The 'predicted' current was the sum of the I_{ATP} and I_{ACh} waveforms; duplicate determinations are shown. Graph shows summary of experiments from 33 cells. **d**, Non-

additivity occurs whether ACh application begins before, at the same time as, or after the start of ATP application. From three cells, I_{ATP} was $-5.6 \pm 0.1 \mu A$, I_{ACh} was $-2.7 \pm 0.3 \mu A$, and $I_{ATP/ACh}$ was $-5.6 \pm 0.2 \mu A$ (middle panel) and $-5.5 \pm 0.2 \mu A$ (right panel); the predicted $I_{ATP/ACh}$ was $-8.3 \pm 0.4 \mu A$, and thus the occluded current equals I_{ACh} .

Table 1 Equal non-additivity between P2X₂ and nicotinic channels under conditions of varying Na⁺ electrochemical gradients

P2X ₂ + α3β4	I_{ATP} (μA)	I_{ACh} (μA)	$I_{ATP/ACh}$ Actual (μA)	Prediction (μA)	Percentage of prediction	<i>n</i>
96 mM Na ⁺ , V_h –60 mV	-6.7 ± 1.1	-7.7 ± 1.2	-7.3 ± 1.2	-14.4 ± 2.1	50.9 ± 2.7	7
20 mM Na ⁺ , V_h –60 mV	-2.5 ± 0.2	-2.1 ± 0.3	-2.2 ± 0.2	-4.6 ± 0.6	49.6 ± 3.0	6
20 mM Na ⁺ , V_h –45 mV	-0.7 ± 0.2	-0.5 ± 0.1	-0.6 ± 0.1	-1.2 ± 0.3	46.4 ± 2.0	6

To reduce the size of the inward currents we replaced extracellular Na⁺ with NMDG⁺ (20 mM Na⁺, 78 mM NMDG⁺). For these experiments, we applied agonist for brief periods only (1–2 s). We observed equal non-additivity with 98 mM Na⁺, 20 mM Na⁺, or 20 mM Na⁺ at a holding potential (V_h) of –45 mV, which is ~5 mV negative to the reversal potential under these conditions.

had a constant Hill slope and half-maximal ATP concentration, regardless of the presence of either $\alpha 3$, $\beta 4$ or both $\alpha 3$ and $\beta 4$ nicotinic subunits ($n = 4-6$; see Supplementary Information, Fig. 1 and Table 1). Similarly, the $\alpha 3\beta 4$ channel ACh concentration-response curve was not affected by expression of the P2X₂ receptor (see Supplementary Information, Table 1). We examined gain-of-function mutants at the 9' position in the M2 region in nicotinic channels¹⁴. The $\alpha 3\text{Ser}9' \rightarrow \text{Leu}$ mutation in the $\alpha 3\beta 4$ channel decreased ACh effective concentration for half-maximum response (EC_{50}) by 375-fold, but this shift in the dose-response relation was not affected by the presence of P2X₂ channels (see Supplementary Information, Table 1). Conversely, the EC_{50} for ATP was similar when cells expressed P2X₂ channels alone, $\alpha 3\beta 4$ channels + P2X₂ channels, or ($\alpha 3\text{Ser}9' \rightarrow \text{Leu}$) $\beta 4$ + P2X₂ channels.

Co-activation of P2X₂ and $\alpha 3\beta 4$ channels elicited currents significantly smaller than the predicted current at concentrations of ACh greater than the EC_{10} . The control ACh EC_{50} value was $20.1 \pm 2.4 \mu\text{M}$, but was $33.0 \pm 6.7 \mu\text{M}$ in the presence of $10 \mu\text{M}$

ATP ($P > 0.05$, $n = 5$; see Supplementary Information, Fig. 1). There was a significant decrease in apparent cooperativity, and the maximum was lower (Fig. 1c; see Supplementary Information, Fig. 1). Hill slopes for ACh concentration-response relations in the absence and presence of ATP ($10 \mu\text{M}$) were 1.4 ± 0.07 and 0.84 ± 0.007 , respectively; these Hill slopes differed significantly at $P = 0.004$ (Supplementary Information, Fig. 1).

The quaternary ammonium local anesthetic derivative QX314 (QX; $100 \mu\text{M}$) is an open-channel blocker of nicotinic channels but not of P2X₂ channels. QX blocked I_{ACh} at $\alpha 3\beta 4$ channels by more than 80% with a time constant of 15 s ($n = 6$; $100 \mu\text{M}$ ACh applied for ~ 75 s; Fig. 2a). If open nicotinic channels contribute to $I_{\text{ATP/ACh}}$, then one expects QX to block $\sim 80\%$ of the nicotinic component with a time constant of 15 s. However, the decay of $I_{\text{ATP/ACh}}$ in the presence of QX was identical to the decay of I_{ATP} (Fig. 2b), indicating that during co-activation substantial proportions of $\alpha 3\beta 4$ nicotinic channels either fail to open, or open rarely. The current-voltage relation of P2X₂ channels is weakly inwardly rectifying, whereas that

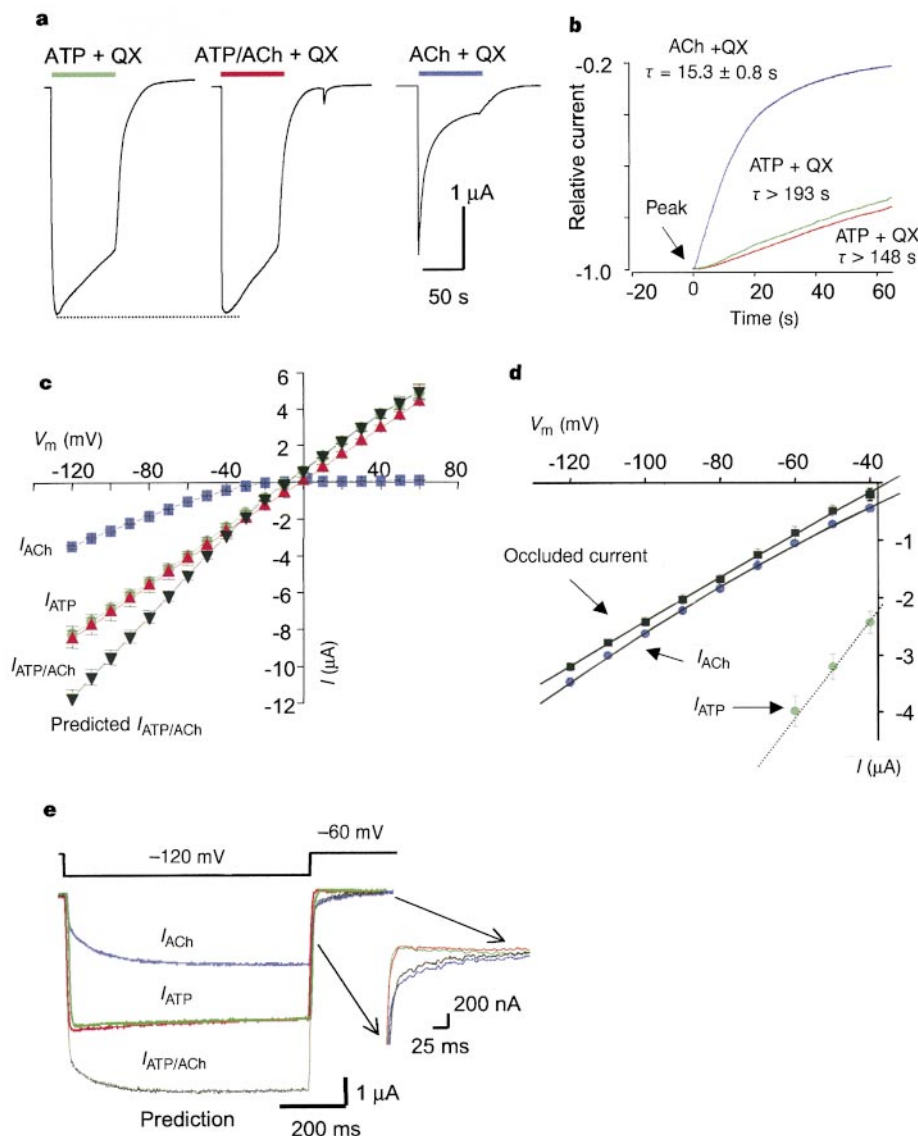


Figure 2 P2X receptor activation results in fewer open nicotinic channels. **a**, Inward currents from *Xenopus* oocytes expressing P2X₂ and $\alpha 3\beta 4$ channels. Currents were evoked with ATP + QX, ATP/ACh + QX, and ACh + QX (all $100 \mu\text{M}$). **b**, Normalized data for the relaxation of the agonist-evoked currents in the presence of QX. Note that the ACh current relaxes exponentially with $\tau = 15$ s, whereas I_{ATP} and $I_{\text{ATP/ACh}}$ relax with $\tau > 148$ s ($n = 6$). **c**, Current-voltage relations for I_{ATP} , I_{ACh} and $I_{\text{ATP/ACh}}$ in cells expressing $\alpha 3\beta 4$ and

P2X₂ channels. **d**, The current-voltage relationship shows the experimentally determined I_{ATP} , I_{ACh} and the occluded current (predicted $I_{\text{ATP/ACh}}$ minus experimentally determined $I_{\text{ATP/ACh}}$). **e**, Traces for voltage-jump relaxations of I_{ATP} ($\tau_1 = 6 \pm 0.2$ ms), I_{ACh} ($\tau_1 = 13 \pm 2$ ms and $\tau_2 = 93 \pm 13$ ms) and $I_{\text{ATP/ACh}}$ ($\tau_1 = 6 \pm 0.1$ ms) in cells expressing $\alpha 3\beta 4$ and P2X₂ channels for a step from -60 to -120 mV and back to -60 mV.

for $\alpha 3\beta 4$ channels is strongly inwardly rectifying¹⁵ (Fig. 2c; see Supplementary Information, Fig. 2). In cells co-expressing P2X₂ and $\alpha 3\beta 4$ channels, $I_{ATP/ACH}$ was the same as I_{ATP} at all membrane potentials where I_{ACH} was greater than zero (Fig. 2c). Subtraction of $I_{ATP/ACH}$ from the predicted $I_{ATP/ACH}$ provides a measure of the 'occluded current' at any voltage. The occluded current superimposed with I_{ACH} at all membrane voltages between -120 and -40 mV ($n = 5$; Fig. 2d).

Because all experiments were done using Ca²⁺-free solutions and because non-additivity was nearly constant between -120 mV and -40 mV, it is unlikely that permeant ions determine non-additivity. But is I_{ATP} additive with I_{ACH} for outward currents? To test this, we expressed muscle $\alpha 1\beta 3\gamma\delta$ nicotinic channels with P2X₂ channels. Evidence for occlusion was obtained in cells where I_{ACH} at -60 mV was at least 50% as great as I_{ATP} . For these cells, $I_{ATP} = -4.4 \pm 0.2 \mu A$; $I_{ACH} = -2.6 \pm 0.1 \mu A$; $I_{ATP/ACH} = -4.6 \pm 0.3 \mu A$; prediction = $-6.7 \pm 0.1 \mu A$ ($n = 6$). Non-additivity for outward currents (see Methods) was slightly less than for inward currents, but occurred to a significant degree (for inward currents $I_{ATP/ACH}$ was $64 \pm 1\%$ of the predicted $I_{ATP/ACH}$, whereas for outward currents $I_{ATP/ACH}$ was $74 \pm 4\%$ of the predicted $I_{ATP/ACH}$, $n = 6$ and 5).

For hyperpolarizing voltage jumps, the $\alpha 3\beta 4$ currents showed bi-exponential current waveforms (Fig. 2e), on a timescale of ~100 ms, to a larger steady-state conductance. For the jump back to -60 mV, the current decreased on a similar timescale¹⁶. In contrast, P2X₂ currents show faster relaxations (< 5 ms for a hyperpolarizing jump) (Fig. 2e). Thus an 'instantaneous' relaxation is produced by P2X₂ channels, whereas a slower relaxation is produced by $\alpha 3\beta 4$ channels. During $I_{ATP/ACH}$ for cells co-expressing P2X₂ and $\alpha 3\beta 4$ channels, both the hyperpolarizing and depolarizing relaxations were dominated by the faster P2X-like components. Note that the 'off-relaxation' to -60 mV resembled the smaller P2X₂ current rather than the larger $\alpha 3\beta 4$ current.

P2X channels increase their permeability to NMDG⁺ in a time-

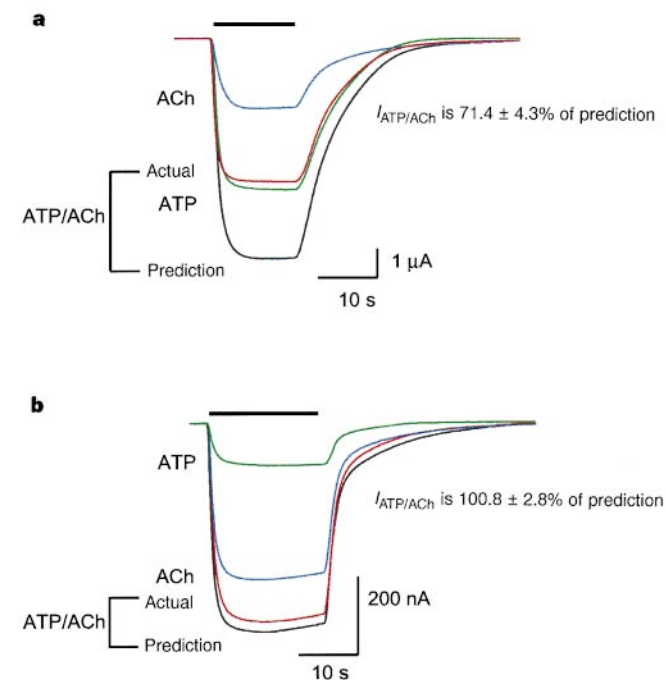


Figure 3 Recovery of P2X₂ and $\alpha 3\beta 4$ channel response additivity. **a**, I_{ATP} , I_{ACH} , $I_{ATP/ACH}$ and the predicted sum of I_{ATP} plus I_{ACH} for cells injected with 1 ng per oocyte of P2X₂, $\alpha 3$ and $\beta 4$ cRNAs. These data are similar to those for 10 ng per oocyte cRNA injections and show clear non-additivity of the type reported in Fig. 1c. **b**, I_{ATP} , I_{ACH} , $I_{ATP/ACH}$ and the predicted sum of I_{ATP} plus I_{ACH} for cells injected with 0.01 ng per oocyte of P2X₂, $\alpha 3$ and $\beta 4$ cRNAs. The ratio of P2X₂: $\alpha 3\beta 4$ was kept constant for all dilution experiments.

and ATP-dependent manner^{17,18}, and the channel reversal potential in NMDG⁺ shifts to positive voltages^{17,18}. In cells co-expressing P2X₂ and $\alpha 3\beta 4$ channels, and activated by ATP, the shift in reversal potential with NMDG⁺ was +18 mV (from -63.8 ± 0.5 to -46.5 ± 2.1 mV; $n = 10$), whereas for cells expressing just P2X₂ channels the shift was +33 mV (from -60.5 ± 0.6 to -27.5 ± 1.5 mV; $n = 10$; see Supplementary Information, Fig. 3). Thus the previously described I₂ state of P2X₂ channels^{17,18} was affected by co-expression of nicotinic channels.

Plasma membrane channel density may be one possible determinant of interactions between glycine-gated channels¹⁹, Kv channels²⁰ and bacterial chemotactic receptors²¹. A stock solution containing P2X₂, $\alpha 3$ and $\beta 4$ cRNAs was diluted at widely varying strengths so that the ratio of P2X₂: $\alpha 3\beta 4$ was held constant, and injected oocytes were assayed after 3 days. Non-additivity occurred equally well for injections of cRNA at 10 ng per oocyte and for 1 ng per oocyte ($n = 4$; Figs 1 and 3a). For injections of cRNA at 0.01 ng per oocyte, however, all responses summed linearly to 100% ($n = 11$), and $I_{ATP/ACH}$ equalled I_{ATP} plus I_{ACH} (Fig. 3b).

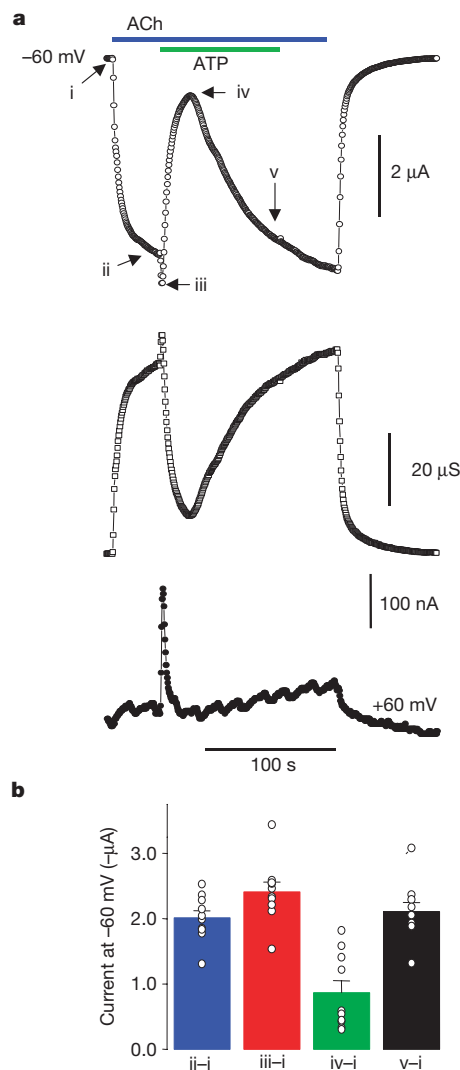


Figure 4 Inhibition of $\alpha 3\beta 4$ channels revealed with a P2X₂ mutant. **a**, Cells co-expressing T18A P2X₂ mutants and $\alpha 3\beta 4$ channels. Upper panel, shows inward current at -60 mV; middle panel, membrane conductance, lower panel; outward current at +60 mV. Cells were activated with ACh followed by ATP/ACh. On application of ATP/ACh, a transient peak is apparent at +60 mV synchronous with the current peak at -60 mV and the peak in membrane conductance (marked iii). **b**, Summary for experiments such as those shown in upper panel of **a**.

Direct proof that nicotinic channels close when P2X₂ channels open would be demonstration of steady-state inhibition of $\alpha 3\beta 4$ channels in real time. This issue was resolved with a P2X₂ channel mutant that opens and desensitizes rapidly—the T18A mutation²². I_{ATP} at T18A P2X₂ mutants desensitized at least 10-fold faster than at wild-type P2X₂ channels (10–90% decay time for T18A P2X₂ was 8.3 ± 2 s; $n = 6$, compared with more than 100 s (refs 11, 17)). We also verified that P2X₂ T18A mutants do not change their ionic selectivity to NMDG⁺ ($n = 6$). As with wild-type channels, $I_{ATP/ACh}$ was smaller than the predicted sum of I_{ATP} and I_{ACh} for T18A P2X₂ mutants and $\alpha 3\beta 4$ channels ($I_{ATP} = -1.9 \pm 0.3$ μ A; $I_{ACh} = -2.0 \pm 0.1$ μ A; $I_{ATP/ACh} = -2.2 \pm 0.2$ μ A; prediction = -4.2 ± 0.4 μ A; $n = 10$). The transient peak (iii in Fig. 4a, upper panel) is dominated by the opening of T18A P2X₂ channels because it is coincident with the P2X₂ current at +60 mV (Fig. 4a, lower panel). T18A channel desensitization reduced the total current to levels lower than I_{ACh} , representing a $60 \pm 9\%$ inhibition of the nicotinic current ($n = 10$; Fig. 4a). After maximal desensitization of T18A P2X₂ mutants (90% after ~ 1 min), I_{ACh} grows back nearly to its value before ATP (growth from iv–v in Fig. 4a, upper panel), with kinetics slower

than those for $\alpha 3\beta 4$ channel activation (growth of ii in Fig. 4a, upper panel).

We co-expressed P2X₄ with $\alpha 3\beta 4$ channels. In the presence of 1 mM Ca²⁺ in the extracellular buffer (to block the P2X₄ I₂ state^{17,18}), we repeated the experiments illustrated in Fig. 4. In cells co-expressing P2X₄ and $\alpha 3\beta 4$ channels, ATP application revealed $40 \pm 2\%$ ($n = 5$) inhibition of the nicotinic current evoked by ACh (at the peak; as in iv of Fig. 4). Thus wild-type P2X receptors produce steady-state inhibition of nicotinic receptors.

We next studied synaptically coupled myenteric neurons from Auerbach's plexus in culture to test for cross-inhibition between P2X and nicotinic channels during cholinergic synaptic transmission¹⁰. Locally applied ATP evoked an inward current in more than 80% of neurons and concomitant with this, the cholinergic excitatory postsynaptic current (EPSC) was occluded at the peak of the ATP-evoked current by $84 \pm 5\%$ within 1 s ($n = 13$; Fig. 5b and c). The time constant for P2X current desensitization was 23 ± 3 s whereas occlusion of EPSCs persisted for at least 4–5 times longer. Thus, even when the ATP-evoked current completely desensitized, the EPSC remained suppressed by $69 \pm 8\%$, implying

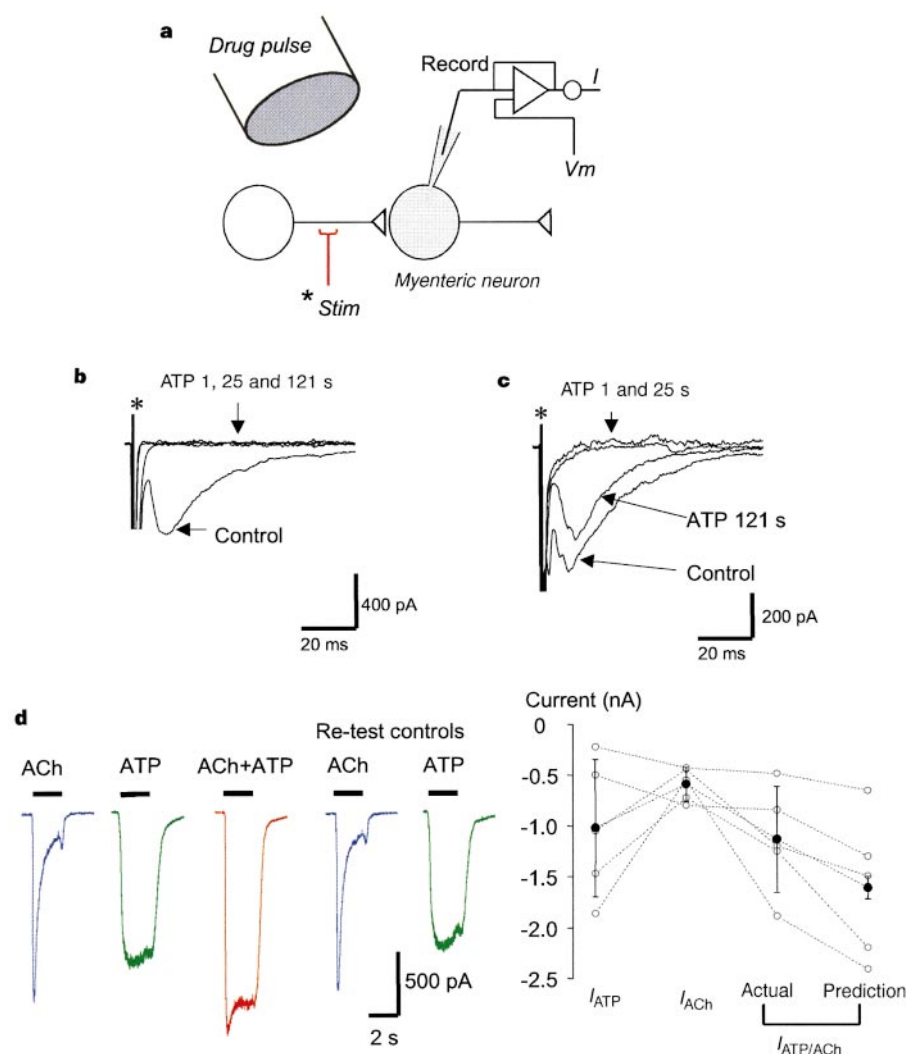


Figure 5 Cross-inhibition between synaptic and somatic nicotinic and P2X channels in synaptically coupled myenteric neurons. **a**, Diagram of recording, stimulating and drug pulsing. **b**, **c**, Nicotinic EPSCs recorded under control conditions and then in the presence of ATP (1 mM) applied for the indicated times (ATP evoked an inward current of -564 ± 374 pA; range -55 to -1205 pA from 13 cells). The EPSC in **b** remains occluded throughout, whereas in **c** it recovers over time; in both cases the steady-state current has

been offset for clarity. **d**, ACh (3 mM) and ATP (1 mM) were pulsed separately to the soma of myenteric neurons, and then both were applied together. Control ACh and ATP pulses were re-applied at the end of the experiment. The scatter graph shows similar experiments (five other cells; note that the predicted $I_{ATP/ACh}$ was larger than the experimentally determined $I_{ATP/ACh}$, $P = 0.02$).

that ACh channels recover slowly from the occluded state. Consistent with our observations on oocytes and synapses, currents evoked by pulsing ACh and ATP onto the soma of myenteric neurons (Fig. 5d) were also non-additive⁷. The occlusion of the cholinergic EPSC by ATP is not mediated by presynaptic or postsynaptic adenosine receptors because occlusion by P2X receptor activation persisted in the presence of the adenosine receptor antagonist 8-cyclopentyltheophylline (CPT; 10 μ M, $n = 14$), whereas the effects of adenosine are blocked. Occlusion of the EPSC occurs only in those neurons that express P2X receptors, and this argues against a role for presynaptic metabotropic P2Y receptors that may affect transmitter release. Moreover, the effect of ATP on the nicotinic EPSC can be mimicked by pulsing ATP and ACh onto the somata of myenteric neurons⁷ (Fig. 5d), and this reproduces the functional interaction with molecularly defined nicotinic and P2X channels. Thus ATP occludes synaptically activated nicotinic channels, and these data may be relevant to the effects of ATP spillover from purinergic synapses¹⁰, ATP released during local tissue damage²³, and ATP released during inflammation²⁴.

Our data indicate that gating transitions in P2X₂ channels are communicated to $\alpha 3\beta 4$ channels and close substantial proportions of them. Cross-inhibition requires both P2X and nicotinic channels, and is not the result of 'co-agonism' or 'co-antagonism' by ATP or ACh at either the nicotinic or P2X channels. Only the I₁ states of P2X₂ channels are detectable in T18A mutants, and these are sufficient to occlude nicotinic channels. During $I_{ATP/ACh}$ in Na⁺ solutions, I_{ATP} sets the peak of the total current, because wild-type P2X₂ channels open at roughly the same rate as nicotinic channels close, and thus steady-state nicotinic current inhibition is not revealed. However, steady-state inhibition is revealed with P2X₂ T18A mutants that lack I₂. Non-additivity is determined at least partially by channel density because decreasing expression prevents occlusion. Ligand bound-open I₁ states of P2X₂ channels cause cross-inhibition of $\alpha 3\beta 4$ nicotinic channels, independently of the direction of current flow. Because we found little or no change in ACh EC₅₀ during cross-inhibition a noncompetitive mechanism for decreased efficacy seems likely. Perhaps the conformational change associated with gating of P2X₂ channels hinders the fenestrations in the inner mouth of nicotinic channels²⁵. Conformational spread from one receptor to its neighbour(s), a concept first identified in bacteria²¹, and recently demonstrated for dopamine D5 receptors and GABA-A channels²⁶, also applies to structurally distinct neuronal transmitter-gated cationic $\alpha 3\beta 4$ nicotinic and P2X₂ channels, in cells and synapses. □

Methods

Heterologous expression

Wild-type P2X₂, P2X₄, $\alpha 3$, $\beta 4$, $\alpha 1$, β , γ and δ cDNAs (all available from previous work^{14,16,17}) were linearized at unique restriction sites downstream of the poly(A) tail. The cDNAs were transcribed *in vitro* using the mMESSAGE mMACHINE kit (Ambion). Each *Xenopus laevis* oocyte was injected with 5–10 ng of cRNA in a volume of 50 nl; the preparation and maintenance of the oocytes was exactly as described¹⁷, and all electrophysiological recordings were made 1–4 days after injection. We carried out site-directed mutagenesis on the cDNAs using synthetic oligonucleotides (Quick Change; Stratagene). Two-electrode voltage-clamp recording of oocytes was performed using the Geneclamp 500 amplifier (Axon Instruments), an Axon digitizing interface and pCLAMP software using described methods and solutions¹⁷. We used NMDG⁺ to measure outward currents through P2X₂ and $\alpha 1\beta\gamma\delta$ channels (about 5 mV positive to the reversal potential) because oocytes could not be held at voltages higher than +20 mV for long periods.

Neurons

Synaptically coupled myenteric neurons were cultured using described techniques¹⁰. Whole-cell patch-clamp recordings were obtained using standard methods at room temperature with patch pipettes with tip resistances of 3–5 M Ω , and filled with a solution comprising (in mM): 160 CsCl, 2 MgCl₂, 1 EGTA, 10 HEPES, 1 ATP and 0.25 GTP. Solutions were adjusted to pH 7.4 (with CsOH) and 315 mosmol kg⁻¹ (with CsCl), respectively¹⁰. Recordings were made with an Axopatch 200A amplifier. Data were acquired with pClamp 8.0 software. Currents were filtered at 1 kHz (4-pole Bessel filter, Warner Instruments, Hamden, Connecticut) and sampled at 5 kHz. Drugs were applied

onto individual neurons by gravity flow from a linear array of quartz tubes (320- μ m i.d. and 450- μ m o.d., Polymicro Technologies). EPSCs were evoked using a pipette stimulating electrode placed on nerve fibres under visual guidance¹⁰.

Data analysis

Data were analysed using Clampfit (Axon Instruments) or Origin 5.0 (Microcal). Data in the text and graphs are shown as mean $\pm$ s.e.m. from n determinations as indicated. All current-voltage relations shown in the figures are leak-subtracted. The ratio of NMDG⁺ to Na⁺ permeability was calculated using the relation $p_{NMDG^+}/p_{Na^+} = \exp(\Delta E_{rev}F/RT)$ where p_{NMDG^+} is the permeability to NMDG⁺, p_{Na^+} is the permeability to Na⁺, ΔE_{rev} is the shift in reversal potential and F , R and T have their standard meanings¹⁷. Concentration-response curves were fitted where appropriate, as indicated, to the Hill equation.

Received 12 April; accepted 26 May 2000.

- Green, T., Heinemann, S. F. & Guse, J. F. Molecular neurobiology and genetics: investigation of neural function and dysfunction. *Neuron* **20**, 427–444 (1998).
- Rogers, M., Colquhoun, L. M., Patrick, J. W. & Dani, J. A. Calcium flux through predominantly independent purinergic ATP and nicotinic acetylcholine receptors. *J. Neurophysiol.* **77**, 1407–1417 (1997).
- Fieber, L. A. & Adams, D. J. Adenosine triphosphate-evoked currents in cultured neurones dissociated from rat parasympathetic cardiac ganglia. *J. Physiol. (Lond.)* **434**, 239–256 (1991).
- Fieber, L. A. & Adams, D. J. Acetylcholine-evoked currents in cultured neurones dissociated from rat parasympathetic cardiac ganglia. *J. Physiol. (Lond.)* **434**, 215–237 (1991).
- Nakazawa, K. ATP-activated current and its interaction with acetylcholine-activated current in rat sympathetic neurons. *J. Neurosci.* **14**, 740–750 (1994).
- Searl, T. J., Redman, R. S. & Silinsky, E. M. Mutual occlusion of P2X ATP receptors and nicotinic receptors on sympathetic neurons of the guinea-pig. *J. Physiol. (Lond.)* **510**, 783–791 (1998).
- Zhou, X. & Galligan, J. J. Non-additive interaction between nicotinic cholinergic and P2X purine receptors in guinea-pig enteric neurons in culture. *J. Physiol. (Lond.)* **513**, 685–697 (1998).
- Barajas-Lopez, C., Espinosa-Luna, R. & Zhu, Y. Functional interactions between nicotinic and P2X channels in short-term cultures of guinea-pig submucosal neurons. *J. Physiol. (Lond.)* **513**, 671–683 (1998).
- Collo, G. *et al.* Cloning of P2X₅ and P2X₆ receptors and the distribution and properties of an extended family of ATP-gated ion channels. *J. Neurosci.* **16**, 2495–2507 (1996).
- Zhou, X. & Galligan, J. J. P2X purinoceptors in cultured myenteric neurons of guinea-pig small intestine. *J. Physiol. (Lond.)* **496**, 719–729 (1996).
- Brake, A. J., Wagenbach, M. J. & Julius, D. New structural motif for ligand-gated ion channels defined by an ionotropic ATP receptor. *Nature* **371**, 519–523 (1994).
- Xu, W. *et al.* Megacystis, mydriasis, and ion channel defect in mice lacking the $\alpha 3$ neuronal nicotinic acetylcholine receptor. *Proc. Natl Acad. Sci. USA* **96**, 5746–5751 (1999).
- Xu, W. *et al.* Multiorgan autonomic dysfunction in mice lacking the $\beta 2$ and $\beta 4$ subunits of neuronal acetylcholine receptors. *J. Neurosci.* **19**, 9298–9305 (1999).
- Labarca, C. *et al.* Channel gating governed symmetrically by conserved leucine residues in the M2 domain of nicotinic receptors. *Nature* **376**, 514–516 (1995).
- Evans, R. J. *et al.* Ionic permeability of, and divalent cation effects on, two ATP-gated cation channels (P2X receptors) expressed in mammalian cells. *J. Physiol. (Lond.)* **497**, 413–422 (1996).
- Figl, A., Labarca, C., Davidson, N., Lester, H. A. & Cohen, B. N. Voltage-jump relaxation kinetics for wild-type and chimeric $\beta 2$ subunits of neuronal nicotinic receptors. *J. Gen. Physiol.* **107**, 369–379 (1996).
- Khakh, B., Bao, X., Labarca, C. & Lester, H. A. Neuronal P2X receptor-transmitter-gated cation channels change their ion selectivity in seconds. *Nature Neurosci.* **2**, 322–330 (1999).
- Virginio, C., MacKenzie, A., Rassendren, F. A., North, R. A. & Surprenant, A. Pore dilation of neuronal P2X receptor channels. *Nature Neurosci.* **2**, 315–321 (1999).
- Taleb, O. & Betz, H. Expression of the human glycine receptor $\alpha 1$ subunit in *Xenopus* oocytes: apparent affinities of agonists increase at high receptor density. *EMBO J.* **13**, 1318–1324 (1994).
- Honore, E. *et al.* Different types of K⁺ channel current are generated by different levels of a single mRNA. *EMBO J.* **11**, 2465–2471 (1992).
- Duke, T. A. & Bray, D. Heightened sensitivity of a lattice of membrane receptors. *Proc. Natl Acad. Sci. USA* **96**, 10104–10108 (1999).
- Boue-Grabot, E., Archambault, V. & Seguela, P. A protein kinase C site highly conserved in P2X subunits controls the desensitization kinetics of P2X(2) ATP-gated channels. *J. Biol. Chem.* **275**, 10190–10195 (2000).
- Lutz, P. L. & Kabler, S. Release of adenosine and ATP in the brain of the freshwater turtle (*Trachemys scripta*) during long-term anoxia. *Brain Res.* **769**, 281–286 (1997).
- Bueno, L. & Fioramonti, J. *Can. J. Gastroenterol.* **13**, 42A–46A (1999).
- Miyazawa, A., Fujiyoshi, Y., Stowell, M. & Unwin, N. Nicotinic acetylcholine receptor at 4.6 Å resolution: transverse tunnels in the channel wall. *J. Mol. Biol.* **288**, 765–786 (1999).
- Liu, F. *et al.* Direct protein-protein coupling enables cross-talk between dopamine D5 and γ -aminobutyric acid A receptors. *Nature* **403**, 274–280 (2000).
- North, R. A. Families of ion channels with two hydrophobic segments. *Curr. Opin. Cell Biol.* **8**, 474–483 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

Thanks to H. Li for assistance with preparation of oocytes, and to other members of the group for comments. A Wellcome Trust (UK) International Prize Travelling Fellowship (to B.S.K.) and the NIH supported this work.

Correspondence and requests for materials should be addressed to B.S.K. (e-mail: balkhakh@its.caltech.edu).

Glycosyltransferase activity of Fringe modulates Notch–Delta interactions

Katja Brückner^{*†}, Lidia Perez^{*}, Henrik Clausen[‡] & Stephen Cohen^{*}

^{*}European Molecular Biology Laboratory, Meyerhofstr 1, 69117 Heidelberg, Germany

[‡]School of Dentistry, University of Copenhagen, Norre Alle 20, DK2200 Copenhagen N, Denmark

Ligands that are capable of activating Notch family receptors are broadly expressed in animal development, but their activity is tightly regulated to allow formation of tissue boundaries¹. Members of the *fringe* gene family have been implicated in limiting Notch activation during boundary formation^{2–8}, but the mechanism of Fringe function has not been determined. Here we present evidence that Fringe acts in the Golgi as a glycosyltransferase enzyme that modifies the epidermal growth factor (EGF) modules of Notch and alters the ability of Notch to bind its ligand Delta. Fringe catalyses the addition of *N*-acetylglucosamine to fucose, which is consistent with a role in the elongation of *O*-linked fucose *O*-glycosylation that is associated with EGF repeats. We suggest that cell-type-specific modification of glycosylation may provide a general mechanism to regulate ligand–receptor interactions *in vivo*.

In the developing *Drosophila* wing, asymmetric activation of Notch by the dorsally expressed ligand Serrate and the ventrally expressed ligand Delta is required to induce Wingless and Vestigial expression and establish a signalling centre at the dorsal–ventral boundary^{9–13}. Fringe is expressed in dorsal cells and contributes to making them more sensitive to Delta and less sensitive to Serrate^{2,4,14}. One means by which Fringe might change the sensitivity of dorsal cells to Notch ligands is by modulating ligand–receptor interaction. Alternatively, Fringe might act indirectly to influence cellular signalling responses to a given level of ligand binding. To distinguish between these possibilities, we measured the effect of Fringe on Notch–Delta binding.

We expressed the extracellular domain of Delta as a secreted alkaline phosphatase fusion protein for use in a ligand-binding assay (Delta–AP; Fig 1a). To measure interaction of Delta–AP with transiently transfected *Drosophila* SL2 cells, bound Delta–AP was quantified using an enzymatic assay for alkaline phosphatase activity. Expression levels of the transfected proteins were monitored by immunoblot analysis in parallel to the binding assays. SL2 cells expressing Notch alone or Fringe alone bound Delta–AP at a very low level (Fig 1b). Co-expression of Notch and Fringe resulted in a large increase in the quantity of bound Delta–AP (Fig 1b), even when the level of Notch expression was lower than in the cells expressing Notch alone. The level of proteolytic processing required for formation of a functional receptor was not increased by co-expression of Fringe. This suggests that Fringe activity may increase the ability of Notch-expressing cells to bind Delta.

Although Fringe and its vertebrate homologues can be found as secreted proteins (Fig 2b; refs 2, 15), genetic analysis has suggested that Fringe acts cell-autonomously in the wing disc^{11,14}. These apparently contradictory observations prompted us to ask whether secreted Fringe can affect Notch–Delta binding. We compared Delta–AP binding to cells co-expressing Notch and Fringe with its binding to Notch-expressing cells that were co-cultured with Fringe-expressing cells for two days before the binding assay (Fig. 1b). Co-cultured cells bound Delta–AP at background levels. Thus, Fringe does not appear to influence the ability of Notch to

bind Delta when provided as an extracellular protein, but does act when co-expressed with Notch.

The requirement for co-expression of Fringe and Notch could be explained if Fringe exerts its activity within the Notch-expressing cell. Fringe and Brainiac have been suggested to show similarity to bacterial glycosyltransferase enzymes¹⁶, and several mammalian glycosyltransferases related to Brainiac have been characterized¹⁷. If Fringe functions as a glycosyltransferase enzyme, it should act in the Golgi. To test this possibility, we prepared a Golgi-tethered version of Fringe in which the first 40 amino acids (including the predicted signal peptide) were replaced by the first 121 amino acids of the Golgi-resident glycosyltransferase GalNAc-T3 (Fig 2a; ref. 18). The resulting fusion protein includes the transmembrane domain of GalNAc-T3, which functions as a Golgi-retention signal¹⁹. Immunoprecipitation from transfected SL2 cells showed that Fringe–GT was expressed at a comparable level to wild-type Fringe, but was not

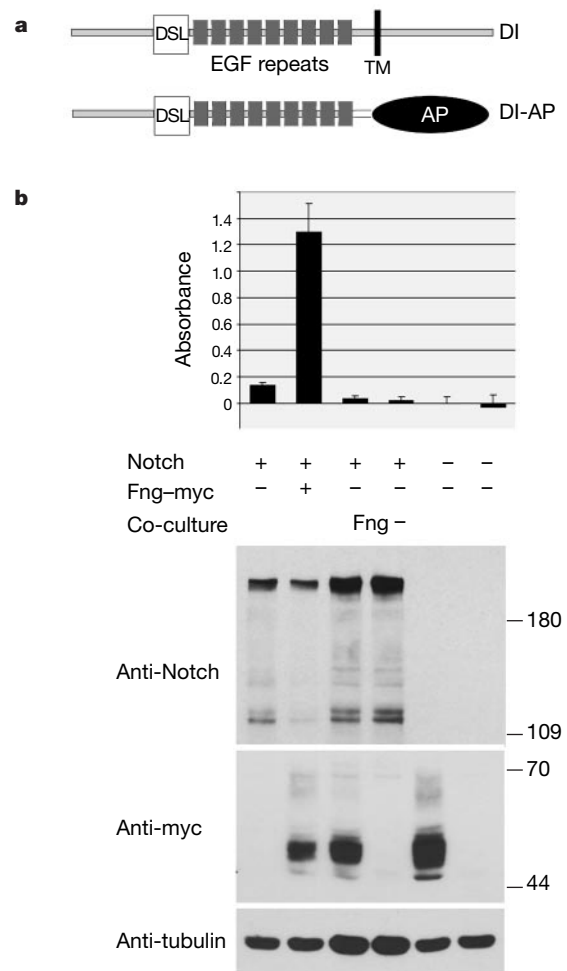


Figure 1 Fringe increases binding of Delta to Notch. **a**, Representation of Delta (DI) and the secreted Delta–alkaline phosphatase fusion protein. TM, transmembrane domain. DSL, Delta Serrate Lag domain. **b**, Upper panel, Delta–AP binding to control and transfected SL2 cells. Cells were transfected with constructs to direct expression of Notch or Fringe–myc as indicated. Cells transfected with empty vector were used as a control. Co-culture indicates that cells transfected with Notch were grown as a mixed culture with cells independently transfected to express Fringe–myc or with cells transfected with empty vector. AP activity is shown in absorbance units. Means $\pm$ s.d. of duplicate binding assays are shown. Lower panels, immunoblots of cell extracts prepared in parallel to those used in the binding assay and probed with anti-Notch, anti-myc and anti-Tubulin. Two forms of Notch are recognized by antibody 9C6. The upper band is the unprocessed form of Notch that does not reach the cell surface²⁴; the lower band is the C-terminal portion of the proteolytically processed form, reflecting production of the mature cell-surface protein.

[†] Present address: Howard Hughes Medical Institute, Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA.)

secreted at detectable levels (Fig. 2b). Immunofluorescent labelling of transfected cells with antibody to a Golgi-resident protein and with anti-myc to visualize Fringe-GT-myc showed that the proteins co-localize (Fig. 2c). Together, these observations confirm that the transmembrane tether provided by GalNAc-T3 is effective in SL2 cells.

In binding experiments, co-expression of Fringe-GT was sufficient to stimulate Delta-AP binding to Notch and was almost as effective as wild-type Fringe (Fig. 2d). This suggests that Fringe-GT has comparable activity to wild-type Fringe. Fringe-GT was also found to be functional *in vivo*, despite not being secreted. When expressed under *patched*^{GAL4} control Fringe-GT had no effect in the dorsal compartment where endogenous Fringe is expressed, but interrupted the endogenous Wingless stripe at the dorsal-ventral boundary and induced ectopic expression of Wingless in the ventral compartment (Fig. 2e, f). These effects are comparable to those caused by expression of wild-type Fringe-myc (Fig. 2g), and suggest that the Golgi-resident form of Fringe has full biological activity. Many Golgi glycosyltransferase enzymes are also found as secreted soluble enzymes, although the function of the secreted forms is unknown. Therefore, the secretion of Fringe proteins may not be of functional significance to their roles as modifiers of Notch activity.

A D-x-D amino-acid motif found in many glycosyltransferases is required for catalytic activity and appears to be involved in coordination of a divalent metal ion in the binding of the donor nucleotide sugar^{16,20,21}. In Fringe, this motif may correspond to residues D236–238. If Fringe acts as a glycosyltransferase, replacing residues 236–238 with asparagine (Fringe-NNN) should destroy enzymatic activity but have a minimal effect on overall protein structure. Consistent with this possibility, co-expression of the Fringe-NNN mutant with Notch did not increase Delta-AP

binding above background levels (Fig. 3a). Furthermore, ectopic expression of Fringe-NNN under *patched*^{GAL4} control did not cause Notch activation in the ventral compartment in the wing imaginal disc (not shown). These observations suggest that Fringe-NNN is inactive *in vivo*.

To determine whether Fringe has intrinsic glycosyltransferase activity, we produced wild-type Fringe and Fringe-NNN by baculovirus infection of insect cells. Microsomal fractions enriched for Golgi membranes were partially solubilized and assayed for the ability of the expressed proteins to catalyse the transfer of ¹⁴C-labelled UDP donor sugars onto acceptor sugars. We tested a variety of different donor-acceptor combinations (Fig. 3b). The highest level of activity was observed with wild-type Fringe microsome lysate and the combination of UDP-N-acetylglucosamine (GlcNAc) and fucose (18-fold over the background level observed with Fringe-NNN). Fringe showed no significant activity with other donor-acceptor combinations.

Fucose is commonly found as an unsubstituted terminal sugar residue in N- and O-linked oligosaccharide chains of glycoproteins and in glycosphingolipids of eukaryotic cells. In contrast, addition of O-linked fucose directly to proteins is a rare type of glycosylation that is found in association with the cysteine-rich consensus sequence C-x-x-G-G-S/T-C (ref. 22). This consensus sequence is found in EGF modules, including a subset of those in Notch, Serrate, Delta, and in their nematode and vertebrate homologues (ref. 23; Fig. 4a). Our results raise the possibility that Fringe functions by elongating O-linked fucose residues in the EGF repeats of Notch through the addition of GlcNAc.

To determine whether Fringe acts through the EGF repeats of Notch, we expressed Notch as a fusion protein in which all sequences following the EGF repeats were replaced by heterologous

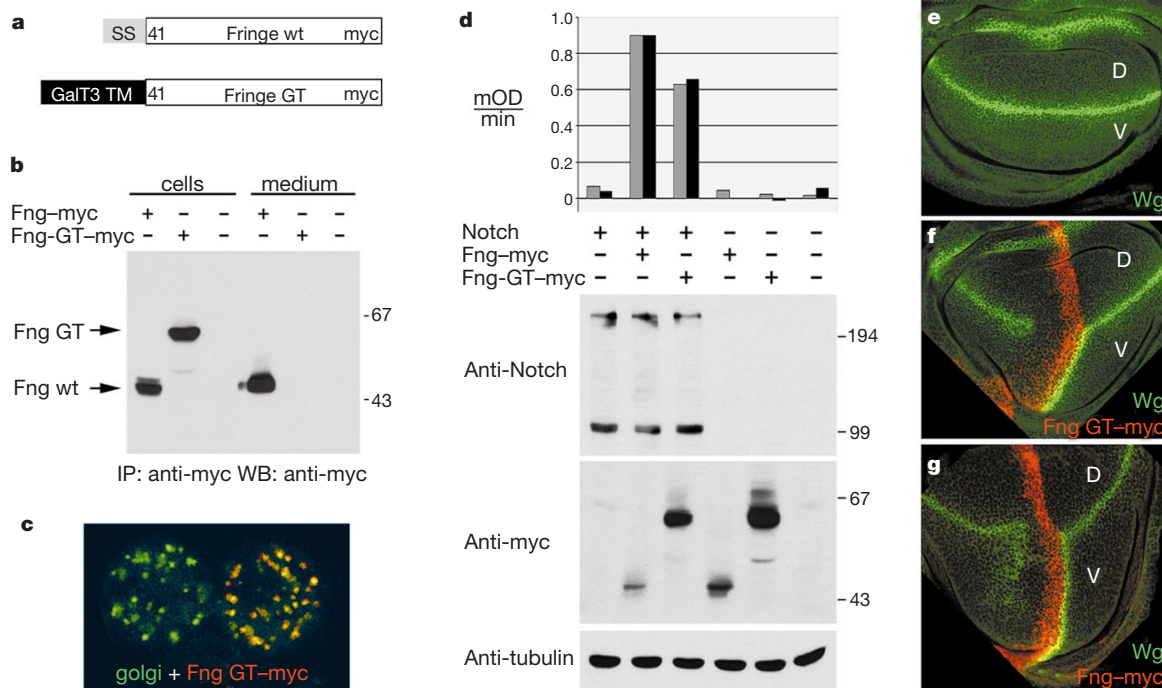


Figure 2 Golgi-tethered Fringe increases binding of Delta to Notch. **a**, Representation of wild-type and Golgi-tethered Fringe (Fng GT). The N-terminal signal sequence (ss) of wild-type Fringe (residues 1–40) was replaced by the transmembrane and juxtamembrane region (1–121) of the Golgi-resident GalNAc-T3 protein. Both proteins carry a C-terminal myc tag. **b**, Western blot of myc-tagged proteins immunoprecipitated from whole cell lysates (cells) and from conditioned medium. Cells were transfected to express wild-type Fringe-myc or Fringe-GT-myc. Control cells were transfected with empty vector. **c**, Immunofluorescent labelling of SL2 cells transfected to express Fringe-GT-myc and labelled with antibody to *Drosophila* Golgi (green) and anti-myc (red). The cell on the right

expressed Fringe-GT-myc. **d**, Delta-AP binding to Notch-expressing cells was increased by co-expression of Fringe-myc or Fringe-GT-myc. Replicate experiments are shown. Immunoblots of cell lysates probed with anti-Notch, anti-myc and anti-Tubulin are shown below. **e–g**, Wing imaginal discs labelled to visualize Wingless protein (green) and the myc epitope tag (red). **e**, *patched*^{GAL4} control wing disc. **f**, *patched*^{GAL4}/UAS-Fringe-GT-myc. **g**, *patched*^{GAL4}/UAS-Fringe-myc. Fringe-GT-myc and wild-type Fringe-myc expression in the *patched*^{GAL4} stripe are shown in red.

sequences from the transmembrane protein CD2 (Fig. 4a). Cells expressing Notch-CD2 and Golgi-tethered Fringe-GT bound over 50-fold more Delta-AP than cells expressing Notch-CD2 alone or control cells (Fig. 4b). This observation indicates that the EGF repeats of Notch are sufficient to mediate binding to Delta when taken out of their normal context. Under normal circumstances, expression of Notch on the cell surface requires proteolytic cleavage in the extracellular domain by a furin-like protease^{24,25}. The cleaved extracellular domain remains attached to the transmembrane and intracellular domain to form an active receptor complex. In the case of Notch-CD2, proteolytic processing does not appear to be

required for cell-surface expression and Delta-AP interaction, as the protein lacks the cleavage site located between Notch-Lin repeats and the transmembrane domain that is used in mouse Notch1.

As the EGF modules of Notch appear to be sufficient to mediate ligand interaction, we reasoned that the corresponding domain of Notch expressed as a soluble AP fusion protein might retain ligand-binding activity (Fig. 4a). We produced Notch-AP in presence or absence of co-expressed Fringe-GT. AP activity of the supernatants was first normalized and binding of the secreted AP fusion proteins was carried out on SL2 cells expressing the full-length form of Delta.

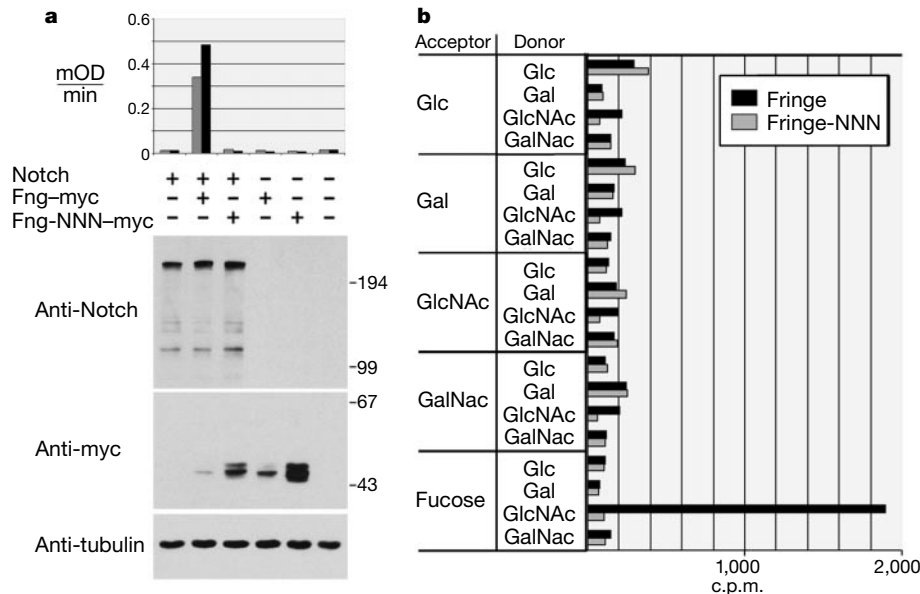


Figure 3 Fringe has glycosyltransferase activity. **a**, Delta-AP binding to control and Notch-transfected SL2 cells. Cells were transfected to express Notch, wild-type Fringe-myc, or mutant Fringe-NNN-myc as indicated. Fringe-NNN-myc has no activity in the binding assay, despite being expressed at higher levels than wild-type Fringe-myc. Lower panels, western blots of cell lysates probed with anti-Notch, with anti-myc and with anti-Tubulin. MOD, absorbance units $\times 10^{-3}$. **b**, Glycosyltransferase activity was

measured in microsomal fractions from cells expressing wild-type Fringe-myc or Fringe-NNN-myc. Enzyme activity was measured by transfer of ^{14}C -labelled sugars from UDP donor sugars onto acceptor sugars. Average results from two experiments are shown. Donors tested were UDP-glucose (Glc), UDP-galactose (Gal), UDP-*N*-acetyl-glucosamine (GlcNAc) and *N*-acetyl-galactosamine (GalNAc). Acceptors tested were D-glucose, D-galactose, D-GlcNAc, D-GalNAc and L-fucose.

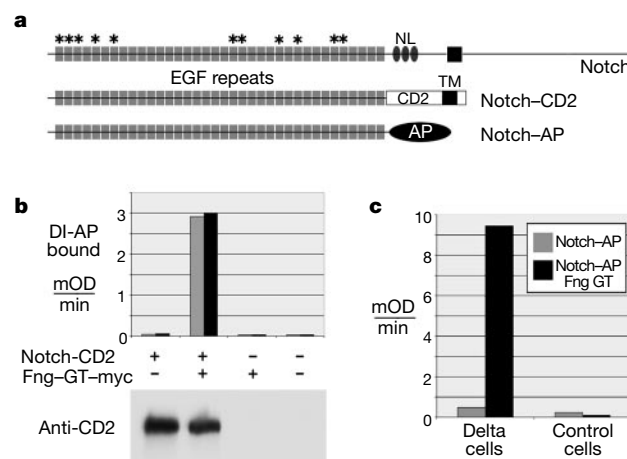


Figure 4 Secreted Notch produced by Fringe-GT-expressing cells binds Delta-expressing cells. **a**, Representation of Notch, Notch-CD2 and the secreted Notch-AP fusion proteins. Notch-CD2 consists of the EGF modules fused to the transmembrane protein CD2. Notch-AP consists of the same EGF modules expressed as a secreted AP fusion protein. Asterisks indicate EGF modules containing a perfect consensus sequence for *O*-linked fucose modification. NL, Notch-Lin repeats; TM, membrane-spanning domain. **b**, Quantitation of Delta-AP binding in replicate experiments. Cells were transfected to express Notch-CD2 or Fringe-GT-myc as indicated. Lower panel,

immunoblot probed with anti-CD2. Notch-CD2 expression was comparable in the presence or absence of Fringe-GT-myc. Notch-CD2 lacks the site for furin-mediated cleavage located near the Notch-Lin repeats²⁵ and migrates at a relative molecular mass of ~200,000. **c**, Quantitation of Notch-AP binding to cells expressing full-length Delta or control cells. Grey bars, Notch-AP produced by SL2 cells; black bars, Notch-AP produced by SL2 cells co-expressing Golgi-tethered Fringe-GT-myc. The level of Notch-AP activity in the conditioned media was normalized.

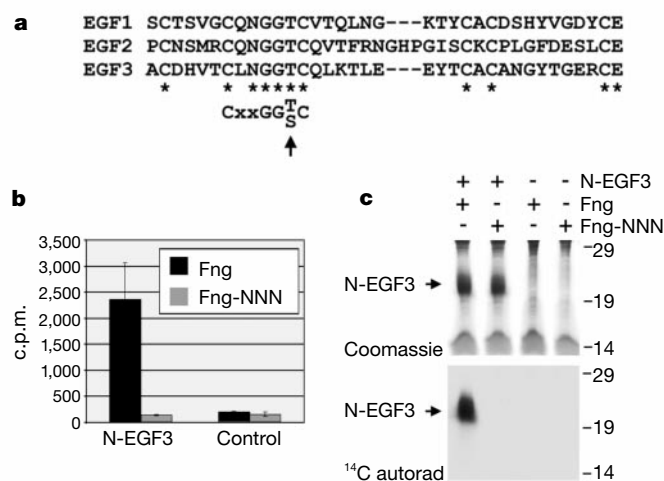


Figure 5 Glycosylation of Notch by Fringe *in vitro*. **a**, Aligned amino-acid sequence of the first three EGF repeats of Notch. Conserved residues and the consensus sequence for *O*-fucosylation are indicated (arrow). **b**, Glycosyltransferase activity of microsomal fractions from cells expressing wild-type Fringe-myc or Fringe-NNN-myc. Enzyme activity was measured by transfer of ¹⁴C-labelled UDP-GlcNAc onto Notch-EGF3-His (N-EGF3). Average results from two experiments are shown. **c**, SDS-PAGE of samples

from **b** run on a 15% acrylamide gel. Upper panel, Coomassie blue stained gel. The Notch-EGF3 protein is indicated. Lower panel, ¹⁴C-UDP-GlcNAc-labelled Notch-EGF3 protein visualized by autoradiography. Background proteins from the microsome fractions were not ¹⁴C-labelled.

Notch-AP produced in cells co-expressing Fringe-GT-myc bound to Delta-expressing cells 20 times more effectively than Notch-AP produced in the absence of Fringe (Fig. 4b). This suggests that the observed binding relies solely on the EGF modules of Notch being present as a secreted soluble protein.

To determine whether Fringe acts directly to modify one or more EGF modules of Notch, we carried out an *in vitro* glycosylation assay using a short protein consisting of the first three EGF modules of Notch as the substrate (EGF3). The first three EGF modules of Notch contain perfect consensus sites for addition of *O*-linked fucose (Fig. 5a). The results presented above suggest that Fringe might act by elongating *O*-linked fucose through addition of GlcNAc. The Notch-EGF3 protein was expressed in SL2 cells with a carboxy-terminal histidine tag to permit purification of the secreted protein. Equal amounts of Notch-EGF3 were incubated with ¹⁴C-labelled UDP-GlcNAc and microsomal lysates from cells expressing wild-type Fringe or Fringe-NNN. Over 10-fold more labelled GlcNAc was incorporated into Notch-EGF3 by wild-type Fringe than by the mutant form Fringe-NNN (Fig. 5b, c). We conclude that Fringe can act directly to modify the EGF repeats of Notch.

Our results provide evidence that Fringe is a glycosyltransferase enzyme that acts in the Golgi to modify Notch. Fringe-dependent glycosylation of Notch increased its ability to bind Delta. Unexpectedly, we were unable to detect measurable binding of Notch-AP to cells expressing Serrate, or of Serrate-AP to cells expressing Notch. Fringe had no measurable effect on Serrate-AP binding to Notch in our assays when expressed in either the Serrate-AP producing cells or in Notch-expressing cells. This suggests that another factor may be required to promote Notch-Serrate binding. *Drosophila* Brainiac shows limited sequence similarity to Fringe¹⁶ and has been implicated as a modulator of the activities of Notch and EGF signalling pathways²⁶. Mammalian Brainiac-related proteins have been characterized as β 3Gal or β 3GlcNAc glycosyltransferases with acceptor substrate specificities distinct from Fringe¹⁷. It remains to be determined whether Brainiac modifies Notch-ligand or other receptor-ligand interactions.

We propose that Fringe activity determines the type of *O*-linked fucose extension on the EGF repeats of Notch, and possibly on other EGF-repeat-containing proteins. *O*-linked fucose may be extended by the addition of β 1-3-glucose or β 1-3GlcNAc, and the latter may be followed by addition of galactose and sialic acid residues^{22,23}. Our

results suggest that Fringe directs elongation of *O*-linked fucose in the EGF modules of Notch by addition of GlcNAc. Fringe-mediated modification changes the properties of Notch-Delta binding and has an important role in conferring signalling specificity *in vivo*. The identification of enzymes that selectively modify oligosaccharide side chains suggests a new range of possibilities for the regulation of ligand-receptor interactions in a cell-type-specific and protein-specific manner. □

Methods

Constructs

Notch-AP was constructed by cloning sequences encoding amino acids 1-1,467 of Notch in frame with human placental alkaline phosphatase from pcDNA3-AP (ref. 27). The fusion junction is located at the *BspEI* site between the last EGF repeat and the first Notch-Lin repeat. The same Notch fragment was used to make Notch-CD2 and was linked in frame to rat CD2 at residue 2. For Notch-EGF3, residues GHHHHHH and a stop codon were introduced after residue 177 of Notch. For Delta-AP, a *BglII* site was introduced by PCR after residue N592 and sequences encoding residues 1-592 of Delta were fused in frame with AP. For Fringe-myc, residues EFEQKLISEEDL were introduced at the C terminus of Fringe by PCR. Fringe-myc was cloned into pRmHa3 for expression in SL2 cells and into pUAST for GAL4-regulated expression in *Drosophila*. For Fringe-NNN-myc, residues D236-D238 of Fringe-myc were converted to N236-N238 by PCR. For Fringe-GT-myc, fragments encoding the first 121 amino acids of GalNAc-T3 (GenBank accession number X92689) and 40-424 of Fringe-myc were amplified by PCR using oligonucleotides that produce a 15-bp overlapping sequence at the fusion junction. The first two PCR products were used as template to amplify the full-length fusion.

Cell culture and binding assays

Complementary DNAs for expression in *Drosophila* Schneider SL2 cells were cloned into pRmHa3. Cells were transiently transfected by the CaPO₄ method using 4-8 μ g of DNA per well in 6-well plates. Expression was induced by addition of 0.7 mM CuSO₄ for 2 days. Conditioned medium was collected from Notch-AP and Delta-AP transfected cells 2-4 days after induction. The activity level of Notch-AP expressed with and without Fringe was normalized by addition of SL2 conditioned medium. AP-containing supernatants were supplemented with 0.1% NaN₃ and incubated with adherent cells for 90 min at room temperature. Cells were washed 5 times with HBSS containing 0.05% BSA and 0.1% azide, and lysed in 10 mM Tris pH 8, 1% Triton-X100. Endogenous AP was inactivated by heat treatment for 10 min at 65 °C and the lysates clarified by centrifugation. AP activity was measured in 1 M diethanolamine, 5 mM MgCl₂, 6.25 mM *p*-nitrophenyl phosphate. Bound AP activity was quantified in 96-well plates using a microplate reader and Micromanager software (BioRad). An additional replicate of each transfection was prepared for immunoblot analysis. Lysates were prepared separately for immunoblot analysis to allow inclusion of protease inhibitors (which were not used in the binding assay).

Glycosyltransferase assays

Fringe-Myc and Fringe-NNN-myc were cloned into baculovirus vector pVL1393

(Pharming) and expressed in High Five cells. Microsomal fractions were prepared by hypotonic lysis followed by ultracentrifugation. Membrane pellets were solubilized 1:2 (vol/vol) in 20 mM cacodylate pH 6.5, 1% Triton-CF54 and 5 mM MnCl₂ containing leupeptin and aprotinin. This suspension (5 µl) was added to a total of 50 µl reaction mixture containing 25 mM cacodylate pH 6.5, 0.25% Triton-CF54, 5 mM MnCl₂, 500 mM free sugar and 100 µM UDP-[¹⁴C]sugar (1,280–2,000 c.p.m. nmol⁻¹). Reactions were incubated at 37 °C for 45–60 min, followed by Dowex-1 anion exchange chromatography and scintillation counting of the flow through²⁸. For *in vitro* glycosylation of Notch–EGF3, we transfected SL2 cells and purified secreted His-tagged Notch–EGF3 from conditioned medium by Ni-NTA affinity chromatography. We carried out *in vitro* glycosylation as described for acceptor sugars using 0.25 µCi [¹⁴C]GlcNAc per reaction. After incubation, the Ni-NTA beads were washed 4 times, and labelled Notch–EGF3 protein was eluted with 250 mM imidazole in SDS–PAGE sample buffer.

Immunoprecipitation and western blots

Cells were lysed in 50 mM Tris pH 7.5, 1% TritonX100, 120 mM NaCl and 30 mM NaF, containing protease inhibitors (see ref. 29). Antibodies for immunoprecipitation and western blots included mouse monoclonal anti-Myc (9E10), rabbit anti-Myc (Santa Cruz Biotechnology), mouse anti-CD2 (Serotec) and mouse anti-Notch 9C6. Mouse anti-Golgi (ref. 30). Protein bands were visualized with peroxidase conjugated secondary antibodies and enhanced chemiluminescence (Amersham).

Received 8 March; accepted 31 May 2000.

1. Irvine, K. D. Fringe, Notch, and making developmental boundaries. *Curr. Opin. Genet. Dev.* **9**, 434–441 (1999).
2. Panin, V. M., Papayannopoulos, V., Wilson, R. & Irvine, K. D. Fringe modulates Notch–ligand interactions. *Nature* **387**, 908–913 (1997).
3. Johnston, S. H. *et al.* A family of mammalian Fringe genes implicated in boundary determination and the Notch pathway. *Development* **124**, 2245–2254 (1997).
4. Fleming, R. J., Gu, Y. & Hukriede, N. A. Serrate-mediated activation of Notch is specifically blocked by the product of the gene fringe in the dorsal compartment of the *Drosophila* wing imaginal disc. *Development* **124**, 2973–2981 (1997).
5. Rodriguez-Esteban, C. *et al.* Radical fringe positions the apical ectodermal ridge at the dorsoventral boundary of the vertebrate limb. *Nature* **386**, 360–366 (1997).
6. Laufer, E. *et al.* Expression of Radical fringe in limb-bud ectoderm regulates apical ectodermal ridge formation. *Nature* **386**, 366–373 (1997).
7. Zhang, N. & Gridley, T. Defects in somite formation in lunatic fringe-deficient mice. *Nature* **394**, 374–377 (1998).
8. Evrard, Y. A., Lun, Y., Aulehla, A., Gan, L. & Johnson, R. L. lunatic fringe is an essential mediator of somite segmentation and patterning. *Nature* **394**, 377–381 (1998).
9. Rulifson, E. J. & Blair, S. S. Notch regulates wingless expression and is not required for reception of the paracrine wingless signal during wing margin neurogenesis in *Drosophila*. *Development* **121**, 2813–2824 (1995).
10. Diaz-Benjumea, F. J. & Cohen, S. M. Serrate signals through Notch to establish a Wingless-dependent organizer at the dorsal/ventral compartment boundary of the *Drosophila* wing. *Development* **121**, 4215–4225 (1995).
11. Kim, J., Irvine, K. D. & Carroll, S. B. Cell recognition, signal induction and symmetrical gene activation at the dorsal/ventral boundary of the developing *Drosophila* wing. *Cell* **82**, 795–802 (1995).
12. Doherty, D., Feng, G., Younger-Shepherd, S., Jan, L.-Y. & Jan, Y.-N. Dorsal and ventral cells respond differently to the Notch ligands Delta and Serrate during *Drosophila* wing development. *Genes Dev.* **10**, 421–434 (1996).
13. de Celis, J. F., Garcia-Bellido, A. & Bray, S. J. Activation and function of Notch at the dorsal-ventral boundary of the wing imaginal disc. *Development* **122**, 359–369 (1996).
14. Irvine, K. & Wieschaus, E. fringe, a boundary specific signalling molecule, mediates interactions between dorsal and ventral cells during *Drosophila* wing development. *Cell* **79**, 595–606 (1994).
15. Wu, J. Y., Wen, L., Zhang, W. J. & Rao, Y. The secreted product of Xenopus gene lunatic Fringe, a vertebrate signaling molecule. *Science* **273**, 355–358 (1996).
16. Yuan, Y. P., Schultz, J., Mlodzik, M. & Bork, P. Secreted fringe-like signaling molecules may be glycosyltransferases. *Cell* **88**, 9–11 (1997).
17. Amado, M., Almeida, R., Schwientek, T. & Clausen, H. Identification and characterization of large galactosyltransferase gene families: galactosyltransferases for all functions. *Biochim. Biophys. Acta* **1473**, 35–53 (1999).
18. Röttger, S. *et al.* Localization of three human polypeptide GalNAc-transferases in HeLa cells suggests initiation of O-linked glycosylation throughout the Golgi apparatus. *J. Cell Sci.* **111**, 45–60 (1998).
19. Nilsson, T. & Warren, G. Retention and retrieval in the endoplasmic reticulum and the Golgi apparatus. *Curr. Opin. Cell Biol.* **6**, 517–521 (1994).
20. Breton, C. & Imberty, A. Structure/function studies of glycosyltransferases. *Curr. Opin. Struct. Biol.* **9**, 563–571 (1999).
21. Gastinel, L. N., Cambillau, C. & Bourne, Y. Crystal structures of the bovine β4galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. *EMBO J.* **18**, 3546–3557 (1999).
22. Harris, R. J. & Spellman, M. W. O-linked fucose and other post-translational modifications unique to EGF modules. *Glycobiology* **3**, 219–224 (1993).
23. Moloney, D. J. & Haltiwanger, R. S. The O-1 fucose glycosylation pathway: identification and characterization of a uridine diphosphoglucose: fucose-β1,3-galactosyltransferase activity from Chinese hamster ovary cells. *Glycobiology* **9**, 679–687 (1999).
24. Blaumueller, C. M., Qi, H., Zagouras, P. & Artavanis-Tsakonas, S. Intracellular cleavage of Notch leads to a heterodimeric receptor on the plasma membrane. *Cell* **90**, 281–291 (1997).
25. Logeat, F. *et al.* The Notch1 receptor is cleaved constitutively by a furin-like convertase. *Proc. Natl Acad. Sci. USA* **95**, 8108–8112 (1998).
26. Goode, S. & Perrimon, N. Brainin and fringe are similar pioneer proteins that impart specificity to notch signaling during *Drosophila* development. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 177–184 (1997).
27. Bergemann, A. D., Cheng, H. J., Brambilla, R., Klein, R. & Flanagan, J. G. ELF-2, a new member of the Eph ligand family, is segmentally expressed in mouse embryos in the region of the hindbrain and newly forming somites. *Mol. Cell. Biol.* **15**, 4921–4929 (1995).
28. Amado, M. *et al.* A family of human β3-galactosyltransferases. Characterization of four members of a UDP-galactose:β-N-acetyl-glucosamine/β-N acetyl-galactosamine β-1,3-galactosyltransferase family. *J. Biol. Chem.* **273**, 12770–12778 (1998).
29. Brückner, K. *et al.* EphrinB ligands recruit GRIP family PDZ adaptor proteins into raft membrane microdomains. *Neuron* **22**, 511–524 (1999).
30. Stanley, H., Botas, J. & Malhotra, V. The mechanism of Golgi segregation during mitosis is cell type-specific. *Proc. Natl Acad. Sci. USA* **94**, 14467–14470 (1997).

Acknowledgements

We thank T. Nilsson for information about Golgi retention sequences; V. Malhotra for antibody to *Drosophila* Golgi; M. Fortini for Notch and Delta expression plasmids; A.-M. Voie for transgenic strains and F. Peverali for his contributions at an early stage of the work. K.B. thanks K. Prydz and D. Toomre for technical discussion; B. Keck and T. Schwientek for introduction to glycosyltransferase assays. H.C. is supported by the Danish Cancer Center and the Velux Foundation.

Correspondence and requests for materials should be addressed to S.C. (e-mail: cohen@embl-heidelberg.de).

Mice overexpressing human uncoupling protein-3 in skeletal muscle are hyperphagic and lean

John C. Clapham*, Jonathan R. S. Arch*, Helen Chapman*, Andrea Haynes*, Carolyn Lister*, Gary B. T. Moore*, Valerie Piercy*, Sabrina A. Carter*, Ines Lehner*, Stephen A. Smith*, Lee J. Beeley†‡, Robert J. Godden§, Nicole Herrity||, Mark Skehel¶, K. Kumar Changani#, Paul D. Hockings#, David G. Reid#, Sarah M. Squires#, Jonathan Hatcher*, Brenda Trail*, Judy Latcham**, Sohaila Rastan††, Alexander J. Harper*, Susana Cadenas‡‡, Julie A. Buckingham‡‡, Martin D. Brand‡‡ & Alejandro Abuin††‡‡

Departments of * Vascular Biology, † Bioinformatics, § Molecular Biology, || Gene Expression Sciences, ¶ Bioanalytical Sciences, ☆ Neurobehavioural Research, and †† Comparative Genetics, SmithKline Beecham Pharmaceuticals, Third Avenue, Harlow, Essex, CM19 5AW, UK
Departments of # Safety Assessment and ** Laboratory Animal Sciences, The Frythe, Welwyn, Hertfordshire, AL6 9AR, UK
‡‡ Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW, UK, and MRC-Dunn Human Nutrition Unit, Hills Road, Cambridge CB2 2XY, UK

Uncoupling protein-3 (UCP-3) is a recently identified member of the mitochondrial transporter superfamily^{1,2} that is expressed predominantly in skeletal muscle^{1,2}. However, its close relative UCP-1 is expressed exclusively in brown adipose tissue, a tissue whose main function is fat combustion and thermogenesis. Studies on the expression of UCP-3 in animals and humans in different physiological situations support a role for UCP-3 in energy balance and lipid metabolism^{3,4}. However, direct evidence for these roles is lacking. Here we describe the creation of transgenic mice that overexpress human UCP-3 in skeletal muscle. These mice are hyperphagic but weigh less than their wild-type littermates. Magnetic resonance imaging shows a striking reduction in adipose tissue mass. The mice also exhibit lower fasting plasma glucose and insulin levels and an increased glucose clearance rate. This provides evidence that skeletal muscle UCP-3 has the potential to influence metabolic rate and glucose homeostasis in the whole animal.

The human α-skeletal actin promoter was used to drive tissue-

‡ Present addresses: Lexicon Genetics, 4000 Research Forest Drive, The Woodlands, Texas 77381, USA (A.A.). Target Genomics, Pfizer Ltd, Sandwich, Kent, CT13 9NJ, UK (L.J.B.).

directed expression of a human UCP-3 transgene in C57BL/6 × CBA mice (Fig. 1a). This promoter is well characterized, and the 2.2-kb fragment used (a gift from R. Prinjha and F. S. Walsh) contains all the necessary elements for selective expression in skeletal muscle⁵. During development, α -cardiac actin is the predominant isoform of sarcomeric α -actin in mice⁶, and it is only post-partum that there is a switch to α -skeletal actin⁵. Thus, by using the α -skeletal actin promoter, any possibility that embryonic expression of UCP-3 might interfere with development was minimized. The UCP-3 overexpressing animals had normal gestation, birth and litter sizes, and they were viable and outwardly healthy. The mice were developmentally normal in that femur length was similar between groups (14.1 ± 0.2 mm in wild-type and 14.2 ± 0.1 mm in UCP-3 transgenic (tg) mice at 17 weeks of age; mean $\pm$ s.e.m. for 10 animals per group). A number of founders were generated and three independent lines were bred to homozygosity. Of these, two independent lines showed a significant reduction in body weight. The third expressed low levels of human UCP-3. The line expressing the highest levels of human UCP-3 was used to examine the phenotype further. Quantitative polymerase chain reaction with reverse transcription (RT-PCR) of messenger RNA for the human transgene in a single UCP-3tg mouse confirmed that expression was largely confined to skeletal muscle with little or no ectopic expression in other tissues (for example, stomach smooth muscle <1% of skeletal muscle) except brown adipose tissue (Fig. 1b). However, measurement of transgenic UCP-3 expression in brown adipose tissue in a larger sample ($n = 10$) showed that the expressed transgene was only 1% of that in skeletal muscle (Table 1). Total UCP-3 expression was increased 66-fold in skeletal muscle but by only 50% in brown adipose tissue (representing 3% of the levels of endogenous UCP-1). The phenotype described here is therefore due primarily to expression of the transgene in skeletal muscle. As expected, protein expression was confined to the mitochondrial fraction of transgenic muscle (Fig. 1c).

Although data on physiological measurements and blood analytes in this report pertain to males, the phenotype was also observed in female UCP-3tg mice. As UCP-3 overexpression in skeletal muscle was proposed to result in an increase in energy expenditure, it was not surprising that the predominant phenotypic characteristic was a reduction in body weight (Fig. 2a). Placing the

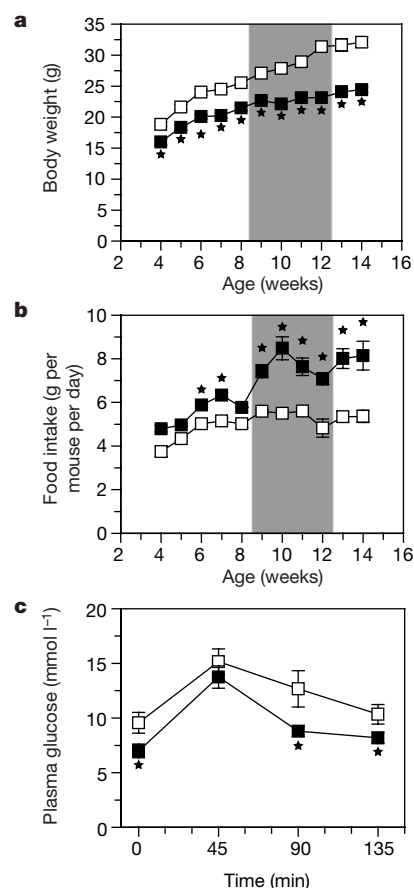


Figure 2 Phenotype of male UCP-3tg mice. **a**, UCP-3tg mice (filled squares) weigh less than their wild-type littermates (open squares) when measured between 4 and 14 weeks of age. Shaded portion of the graph indicates a switch to palatable diet and coincides with a plateau in weight gain. **b**, Twenty-four-hour food intake in grams per mouse, other indices as for **a**. **c**, Oral glucose tolerance curve at the end of the initial period on normal diet (eight weeks). Data are mean $\pm$ s.e.m. (on some data points error bars are within symbols; asterisk, $P < 0.05$) for 12 animals per group.

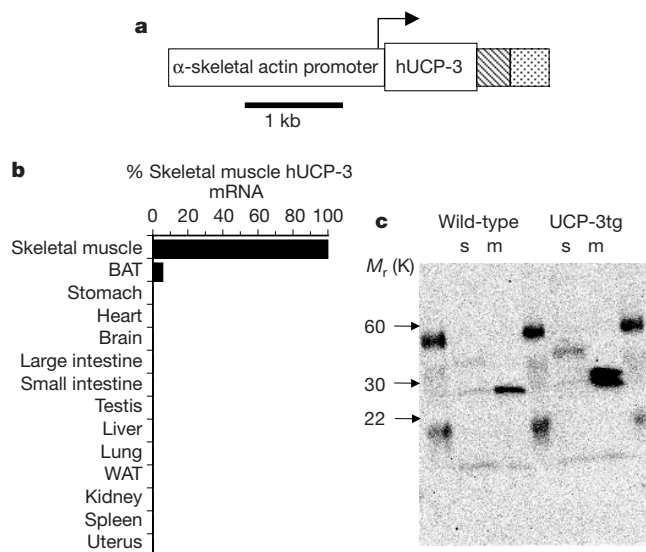


Figure 1 Expression of human UCP-3 in mouse skeletal muscle. **a**, Map of the 3.9-kilobase (kb) construct used for microinjection of [C57BL/6 × CBA] F2 fertilized eggs. Hatched and stippled blocks: artificial intron and SV40 polyadenylation signal from pIRES plasmid, respectively. **b**, Quantitative RT-PCR (TaqMan, PE Biosystems) for human UCP-

3. BAT and WAT, brown and white adipose tissue, respectively. **c**, Immunoblot of human UCP-3 expression at relative molecular mass 34,000 (M_r 34K) in the mitochondrial fraction (m) of UCP-3tg mice, but not wild type, using a rabbit anti-human UCP-3 antibody (Alpha Diagnostic Inc.). s, supernatant.

animals on an energetically equivalent, but palatable diet⁷ encouraged consumption (for example, 96.5 kJ per 24 h compared with 78.6 kJ per 24 h for UCP-3tg mice). It also accelerated weight gain in wild-type animals, but not in UCP-3tg animals: weight gain in UCP-3tg mice tended to plateau at the point of diet change. A surprising finding was that, in spite of their markedly lower body weight, UCP-3tg mice were hyperphagic. They consumed between 15% and 28% more food energy than wild-type mice on the normal diet from 4 to 8 weeks of age, and 33–54% more on the palatable diet between 8 and 12 weeks of age (Fig. 2b). Palatable diets evoke increased sympathetic nerve activity^{8–10}, which stimulates thermogenesis in brown adipose tissue. This, together with the effects of UCP-3 overexpression, may account for the plateau of body weight gain despite increased energy intake. Moreover, further activation of transgenic UCP-3 by products of sympathetically mediated lipolysis (for example, fatty acids) cannot be ruled out. The difference in energy intake between transgenic mice and their wild-type littermates was maintained when the animals were returned to a normal diet. Despite a 50% increase in food consumption, plasma triglycerides and non-esterified fatty acids were similar between the UCP-3tg mice and wild-type controls, suggesting that fat combustion was higher in UCP-3tg mice (Table 1).

Magnetic resonance imaging (MRI) analysis following four weeks on the palatable diet revealed a striking reduction in adipose tissue mass in UCP-3tg mice (Fig. 3a). A 44% and 57% decrease in the ratio of adipose tissue volume to total animal volume was seen in males and females, respectively (Fig. 3b). These alterations may be the predominant factor contributing to reduced weight in these animals. Despite the reduction in adipose tissue content of the UCP-3tg mice, plasma leptin levels were not significantly reduced (Table 1).

The phenotype of the UCP-3tg mouse is consistent with an increase in energy expenditure, which was confirmed directly. Resting oxygen consumption was 25% higher with normal diet at 8 weeks of age (30.5 ± 2.2 ml per animal per hour versus 38.3 ± 1.6 ml per animal per hour in UCP-3tg mice; $P < 0.05$) and 40% with palatable diet at 12 weeks of age (39.8 ± 3.0 ml per animal per hour in wild-type versus 55.9 ± 4.0 ml per animal per hour in UCP-3tg mice; $P = 0.03$). However, locomotor activity was not significantly increased (Table 1). If oxygen consumption is corrected by body weight, the increased consumption of the transgenic mice was 77% ($P < 0.02$) on normal diet and 91% ($P < 0.005$) on palatable diet. Time spent on an accelerating Rotorod, as an index of muscle motor coordination, was also unaffected by UCP-3

overexpression (Table 1). Core temperature was not affected by the presence of the transgene ($38.32 \pm 0.07^\circ\text{C}$ in wild-type and $38.33 \pm 0.13^\circ\text{C}$ in UCP-3tg mice), although muscle temperature was increased ($37.52 \pm 0.32^\circ\text{C}$ in wild-type mice versus $38.72 \pm 0.38^\circ\text{C}$ in UCP-3tg; $P < 0.05$) at 14 weeks of age. The increased muscle temperature was similar to the increase in brown adipose tissue temperature elicited by β_3 -adrenoceptor agonists in rats¹¹. Rodents can sustain marked increases in metabolic rate, with little increase in the temperature of the heat-generating tissue, by rapidly dissipating heat to their environment by radiation and convection¹².

An emerging hypothesis on the role of UCP-3 favours a primary role in lipid substrate utilization^{3,4}. This model was proposed because UCP-3 mRNA levels change markedly, in parallel with altered fatty acid oxidation. For example, 24-hour fasting increases UCP-3 mRNA 6–12-fold, depending on muscle type, without overt changes in whole-body metabolic rate^{3,4}. Thermogenesis is clearly occurring in UCP-3tg mice, but in this context it may be a by-product of greatly increased fatty acid oxidation. Similarly, it has been argued that UCP-1-mediated thermogenesis may be a by-product of fatty acid oxidation, so that animals can survive on a low protein diet without becoming obese¹³.

UCP-3tg mice also displayed reduced fasting plasma glucose levels, increased glucose clearance following an oral glucose load and reduced plasma insulin levels (Fig. 2c and Table 1), indicating that they may be more insulin-sensitive than their wild-type littermates. β_3 -adrenoceptor agonists, which increase metabolic rate and thermogenesis, also elicit weight loss in rats by reducing white adipose tissue mass, and increase insulin sensitivity^{14–16}.

We measured the respiration rate and membrane potential of isolated skeletal muscle mitochondria¹⁷ in six independent paired experiments. UCP-3 overexpression was associated with

Table 1 Summary of biological data

	Wild type	UCP-3tg	P
mRNA expression (cDNA units per ng total RNA $\times 10^3$)			
Skeletal muscle			
Endogenous UCP-3	1.15 ± 0.13	0.93 ± 0.10	ns
hUCP-3 transgene	–	74.4 ± 8.25	
Brown adipose tissue			
Endogenous UCP-1	26.0 ± 1.29	28.4 ± 2.40	ns
Endogenous UCP-3	1.63 ± 0.10	1.68 ± 0.13	ns
hUCP-3 transgene	–	0.85 ± 0.28	
Five-hour-fasted plasma lipid and hormone levels			
Triglycerides (mmol l ⁻¹)	1.70 ± 0.08	2.02 ± 0.19	ns
Non-esterified fatty acids (mmol l ⁻¹)	1.22 ± 0.06	1.19 ± 0.1	ns
Total cholesterol (mmol l ⁻¹)	4.63 ± 0.13	2.90 ± 0.17	<0.001
Leptin (ng ml ⁻¹)	4.50 ± 1.33	3.72 ± 0.84	ns
Insulin (ng ml ⁻¹)	6.16 ± 1.71	1.78 ± 0.34	<0.02
Activity measurements			
Locomotor activity (total 24 h)	$26,233 \pm 1,796$	$33,677 \pm 2,940$	ns
Rotorod (s)	183 ± 27	196 ± 17	ns

Data are expressed as mean $\pm$ s.e.m. for 10–12 animals per group. Plasma measurements and total RNA extractions were from terminal blood and tissue sampling at 19 weeks. UCP mRNA levels were determined using quantitative RT-PCR (TaqMan). Standard curves were constructed using cDNA for each UCP. mRNA expression is represented as cDNA units per ng total RNA. Locomotor activity was conducted in unhabituated animals and the data represent the sum of all transits over a 24-h period as described in ref. 26. *P* values were calculated by Student's unpaired *t*-test; ns, not significantly different from wild type ($P > 0.05$).

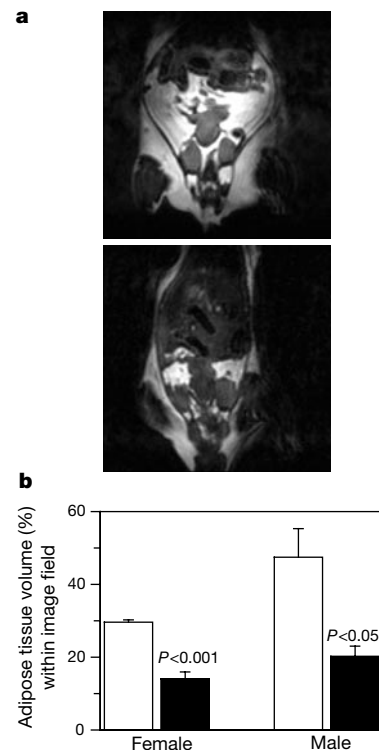


Figure 3 MRI analysis of UCP-3tg mice. **a**, Abdominal images of wild-type and UCP-3tg mice following four weeks on palatable diet. **b**, Comparison between fat as a percentage of body volume in wild-type and UCP-3tg mice. Seven two-dimensional transverse sections for the anterior end of the left kidney through to the posterior end of the right kidney were used for the calculation. Graphs represent mean values of % fat to body volume $\pm$ s.e.m. for six animals per group. Statistical significance was calculated using Student's *t*-test.

mitochondrial uncoupling, measured as a decrease in respiratory control ratio from 3.4 ± 0.1 to 2.4 ± 0.1 ($P < 0.001$). This was caused by a $26 \pm 9\%$ increase in state four respiration (using oligomycin to inhibit ATP synthesis, $P < 0.003$) from a control value of 109 ± 6 nmol O_2 per minute per milligram protein, and by a $12 \pm 6\%$ decrease in state three respiration ($P = 0.04$). In parallel experiments (using nigericin to allow exchange of protons with potassium ions), the mitochondrial membrane potential of 168 ± 5 mV decreased by $12 \pm 3\%$ ($P = 0.001$). Increased state four respiration accompanied by a decreased membrane potential is diagnostic of mitochondrial uncoupling. However, changes in mitochondrial proton conductance following physiologically mediated changes in UCP-3 levels have not been described, and a mitochondrial transport function of UCP-3 cannot be ruled out.

Some studies indicate a role for UCP-3 in the regulation of energy expenditure in humans^{18,19}, but there are also studies that fail to make an association²⁰. The data here are clear, however: despite increased energy intake, overexpression of UCP-3 elicits a marked reduction in body weight, a marked reduction in adiposity and improved insulin sensitivity. Our data are not evidence that UCP-3 is involved in the aetiology of obesity, but they do suggest that enhancement of UCP-3 expression, or stimulation of its activity, is a promising approach for the treatment of this disease. □

Methods

Animals

Mice were maintained in accordance with procedures outlined in the Home Office Animals (Scientific procedures) Act 1986. They were housed in groups of 12 (in cages of 3) and maintained on a 12-h light/12-hour dark cycle (lights on at 06:00 GMT). Normal diet was Teklad 2018 (13.7 kJ g⁻¹) and the palatable diet (12.6 kJ g⁻¹) was prepared as described⁷. The animals were implanted with Implantable Programmable Temperature Transponder chips (IPTT-100, Biomedic Data Systems Inc.) over the rear thigh muscles for identification and muscle temperature estimation.

Generating the transgenic mouse

The transgene fragment (Fig. 1a) was excised from its vector by restriction digestion and purified for injection (2 ng μ l⁻¹). Donor mice were prepared using standard procedures²¹. A Zeiss Axiovert 135M inverted microscope and Leica micromanipulators were used for the microinjection of fertilized eggs. The animals were maintained in a VAF+ animal house. The recipient eggs were [C57BL/6 $\times$ CBA] F2 hybrids.

RNA analysis and immunoblotting

Quantification of mRNA transcripts was performed using the PCR-based 5' nuclease assay²² with gene-specific fluorogenic TaqMan hydrolysis probes²³. Total RNA was treated with deoxyribonuclease I (Promega) and reverse transcribed using random nonamers (Stratagene) and MMLV reverse transcriptase (Life Technologies). After first-strand cDNA synthesis, transcript cDNAs were measured using TaqMan assay oligonucleotide primers and fluorogenic probes designed for human and mouse UCP-3 (amplicon 915–1011 of accession U84763 and amplicon 51–128 of accession AF032902, respectively) and mouse UCP-1 (amplicon 11–109 of accession U63419).

Immunoblots of human UCP-3 were carried out on mouse mitochondrial fractions²⁴ and separated on SDS–PAGE using rabbit anti-human UCP-3 antibodies (Alpha Diagnostic Inc.), anti-rabbit IgG horse radish peroxidase secondary antibody (Amersham) and ECL detection kit (Pierce-Warriner).

Blood analyses

We measured leptin and insulin in plasma taken from 5-h fasted mice using commercially available ELISA (Crystal Chem Inc.). Plasma glucose, triglycerides, non-esterified fatty acids and total cholesterol were measured in the same samples spectrophotometrically using a Cobas Mira plus Clinical Chemistry Analyser (Roche Diagnostics).

Physiological measurements

For food intake measurements, mice were transferred to cages with grid bottoms for 24 h once a week. Body weight was also determined weekly. Oral glucose tolerance tests were performed at 8 weeks of age according to ref. 25. Oxygen consumption was measured using a Servomax oxygen analyser model 580A. Samples were monitored continuously in 20-min blocks with room air as reference. Mice were placed in the chamber during the light phase and the data collected for a 5-h period after the animals had settled into the new environment, but before activity increased owing to anticipation of the dark phase.

Magnetic resonance imaging analysis

Images were obtained on a Bruker AMX300 interfaced to a 18.3-cm 7-T magnet. Animals were anaesthetized using 5% isoflurane and maintained using 1.5% isoflurane and 100%

O_2 . We monitored heart rate and respiration using an ECG and a fibre-optic mechanical probe, respectively. Data acquisitions were triggered to the flat part of the respiratory cycle. A three-dimensional multi-echo sequence was used, acquiring a field of view of $6 \times 6 \times 6$ cm and a matrix size of $128 \times 96 \times 96$. The experimental time was 38 min using 8 echoes with an echo time of 6.6 ms, a repetition time of ~ 2 s and a spectral width of 100 kHz. The data set was zero filled to $256 \times 128 \times 128$ for reconstruction.

All three-dimensional images were quantified using seven sequential transverse sections, 2 mm apart. The total area quantified and the anatomical boundaries were consistent between each animal and the fat depots within each mouse were quantified using the standard software tools available on ParaVision 2.0 (Bruker).

Activity measurements

Locomotor activity was tested in unhabituated mice as described²⁶. Mice (one per cage) were placed in the activity monitor for 24 h, with food and water available *ad lib*.

Locomotor activity was recorded automatically every 30 min. Rotorod experiments were conducted as described²⁷.

Received 24 January; accepted 8 May 2000.

- Boss, O. *et al.* Uncoupling protein-3, a new member of the mitochondrial carrier family with tissue-specific expression. *FEBS Lett.* **408**, 39–42 (1999).
- Vidal-Puig, A., Solanes, G., Gruijic, D., Flier, J. S. & Lowell, B. B. UCP3—an uncoupling protein homolog expressed preferentially and abundantly in skeletal-muscle and brown adipose-tissue. *Biochem. Biophys. Res. Commun.* **235**, 79–82 (1997).
- Samec, S., Seydoux, J., Dulloo, A. G. Role of UCP homologs in skeletal-muscles and brown adipose-tissue—mediators of thermogenesis or regulators of lipids as fuel substrate. *FASEB J.* **12**, 715–724 (1998).
- Samec, S., Seydoux, J., Dulloo, A. G. Post-starvation gene expression of skeletal muscle uncoupling protein 2 and uncoupling protein 3 in response to dietary fat levels and fatty acid composition—A link with insulin resistance. *Diabetes* **48**, 436–441 (1999).
- Brennan, K. J. & Hardeman, E. C. Quantitative analysis of the human alpha-skeletal actin gene in transgenic mice. *J. Biol. Chem.* **268**, 719–725 (1993).
- Gunning, P. *et al.* Differential patterns of transcript accumulation during human myogenesis. *Mol. Cell. Biol.* **7**, 4100–4114 (1987).
- Widdowson, P. S., Upton, R., Buckingham, R., Arch, J. & Williams, G. Inhibition of food response to intracerebroventricular injection of leptin is attenuated in rats with diet-induced obesity. *Diabetes* **46**, 1782–1785 (1997).
- Bellisle, F., Louis-Sylvestre, J., Demozay, F., Blazy, D. & Le Magnen, J. Cephalic phase of insulin secretion and food stimulation in humans, a new perspective. *Am. J. Physiol.* **249**, E639–E645 (1985).
- Leblanc, J. & Brondel, L. Role of palatability on meal-induced thermogenesis in human subjects. *Am. J. Physiol.* **248**, E333–E336 (1985).
- Diamond, P., Brodel, L. & LeBlanc, J. Palatability and postprandial thermogenesis in dogs. *Am. J. Physiol.* **248**, E75–E79 (1985).
- Arch, J. R. *et al.* Treatment of obesity with thermogenic beta-adrenoceptor agonists, studies on BRL 26830A in rodents. *Int. J. Obes.* **8** (suppl. 1), 1–11 (1984).
- Gordon, C. J. *Temperature Regulation in Laboratory Rodents* (Cambridge Univ. Press, 1993).
- Stock, M. J. Gluttony and thermogenesis revisited. *Int. J. Obes.* **23**, 1105–1117 (1999).
- Largis, E. E., Burns, M. G., Muenkel, H. A., Dolan, J. A. & Claus, T. H. Antidiabetic and antiobesity effects of a highly selective β_3 -adrenoceptor agonist (CL 316,243). *Drug Dev. Res.* **32**, 69–76 (1994).
- Weyer, C., Tataranni, A., Snitker, S., Danforth, E. & Ravussin, E. Increase in insulin action and fat oxidation after treatment with CL 316,243, a highly selective β_3 -adrenoceptor agonist in humans. *Diabetes* **47**, 1555–1561 (1998).
- Liu, X., Perusse, F. & Bukowski, L. J. Mechanisms of the antidiabetic effects of the beta 3-adrenergic agonist CL-316243 in obese Zucker-ZDF rats. *Am. J. Physiol.* **274**, R1212–R1219 (1998).
- Cadenas, S. *et al.* UCP2 and UCP3 rise in starved rat skeletal muscle but mitochondrial proton conductance is unchanged. *FEBS Lett.* **462**, 257–260 (1999).
- Argyropoulos, G. *et al.* Effects of mutations in the human uncoupling protein 3 gene on the respiratory quotient and fat oxidation in severe obesity and type 2 diabetes. *J. Clin. Invest.* **102**, 1345–1351 (1998).
- Otobe, S. *et al.* A genetic variation in the 5' flanking region of the UCP3 gene is associated with body mass index in humans in interaction with physical activity. *Diabetologia* **43**, 245–249 (2000).
- Chung, W. K. *et al.* The long isoform uncoupling protein-3 (UCP3L) in human energy homeostasis. *Int. J. Obes.* **23**, S49–S50 (1999).
- Hogan, B., Lacy, E., Beddington, R. & Constantini, F. *Manipulating the Mouse Embryo* 2nd edn (Cold Spring Harbor Press, 1994).
- Holland, P. M., Abramson, R. D., Watson, R. & Gelfand, D. H. Detection of specific polymerase chain reaction product by utilizing the 5'-3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. Natl Acad. Sci. USA* **88**, 7276–7280 (1991).
- Livak, K. J., Flood, S. J., Marmaro, J., Giusti, W. & Deetz, K. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl.* **4**, 357–362 (1995).
- Ernster, L. & Nordenbrand, K. Skeletal muscle mitochondria. *Methods Enzymol.* **10**, 86–94 (1967).
- Young, P. W. *et al.* Repeat treatment of obese mice with BRL 49653, a new potent insulin sensitizer, enhances insulin action in white adipocytes. Association with increased insulin binding and cell-surface GLUT4 as measured by photoaffinity labeling. *Diabetes* **44**, 1087–1092 (1995).
- Rogers, D. C. *et al.* Use of SHIRPA and discriminant analysis to characterise marked differences in the behavioural phenotype of six inbred mouse strains. *Behav. Brain. Res.* **105**, 207–217 (1999).
- Jones, B. J. & Roberts, D. J. The quantitative measurement of motor inco-ordination in naive mice using an accelerating rotarod. *J. Pharm. Pharmacol.* **20**, 302–304 (1968).

Acknowledgements

S.C. was supported by a Marie Curie (EC) fellowship.

Correspondence and requests for materials should be addressed to J.C.C. (e-mail: John_Clapham-1@sbphrd.com).

GATA3 haplo-insufficiency causes human HDR syndrome

Hilde Van Esch^{*†}, Peter Groenen^{*†}, M. Andrew Nesbitt[‡], Simone Schuffenhauer[§], Peter Lichtner[§], Gert Vanderlinden^{*}, Brian Harding[‡], Rolf Beetz^{||}, Rudolf W. Bilous[¶], Ian Holdaway[#], Nicholas J. Shaw[‡], Jean-Pierre Fryns^{**}, Wim Van de Ven^{*}, Rajesh V. Thakker[‡] & Koenraad Devriendt^{**}

^{*} Laboratory for Molecular Oncology, Centre for Human Genetics, University of Leuven and Flanders Interuniversity Institute for Biotechnology, Herestraat 49, B-3000, Leuven, Belgium

[‡] Molecular Endocrinology Group, Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

[§] Department of Medical Genetics, Children's Hospital, Ludwig-Maximilians-University, Goethestrasse 29, 80336 Munich, Germany

^{||} University Children's Hospital, Langenbeckstrasse 1, 55101, Mainz, Germany

[¶] Department of Medicine, University of Newcastle upon Tyne, Framlington Place, Newcastle upon Tyne NE2 4HH, UK

[#] Department of Endocrinology, Auckland Hospital, Park Road, Auckland 1, New Zealand

[‡] Department of Endocrinology, Birmingham Children's Hospital, Steelhouse Lane, Birmingham B46 1YH, UK

^{**} Department of Clinical Genetics, Centre for Human Genetics, University of Leuven, Herestraat 49, B-3000 Leuven, Belgium

[†] These authors contributed equally to this work

Terminal deletions of chromosome 10p result in a DiGeorge-like phenotype that includes hypoparathyroidism, heart defects, immune deficiency, deafness and renal malformations¹. Studies in patients with 10p deletions have defined two non-overlapping regions that contribute to this complex phenotype. These are the DiGeorge critical region II (refs 1, 2), which is located on 10p13-14, and the region for the hypoparathyroidism, sensorineural deafness, renal anomaly (HDR) syndrome³ (Mendelian Inheritance in Man number 146255)⁴, which is located more telomeric (10p14-10pter)^{5,6}. We have performed deletion-mapping studies in two HDR patients, and here we define a critical 200-kilobase region which contains the *GATA3* gene⁷. This gene belongs to a family of zinc-finger transcription factors that are involved in vertebrate embryonic development⁸⁻¹⁰. Investigation for *GATA3* mutations in three other HDR probands identified one nonsense mutation and two intragenic deletions that predicted a loss of function, as confirmed by absence of DNA binding by the mutant *GATA3* protein. These results show that *GATA3* is essential in the embryonic development of the parathyroids, auditory system and

kidneys, and indicate that other *GATA* family members may be involved in the aetiology of human malformations.

To delineate the region containing the HDR locus, we characterized a complex reciprocal insertional translocation with a breakpoint at 10p15 that had been previously reported in an HDR patient⁵ (5/99, Table 1). As a first step, we constructed a 1.2-megabase (Mb) contig of the breakpoint region (Fig. 1a), which consisted of yeast artificial chromosome (YAC), bacterial artificial chromosome (BAC) and P1-derived artificial chromosome (PAC) clones, and used these in fluorescence *in situ* hybridization (FISH) experiments to define a deletion of about 900 kilobases (kb) (Fig. 1a, b) in this HDR patient. Segregation analysis in other HDR patients from family 23/91 (ref. 3), using the polymorphic microsatellite markers D10S189, D10S1751, D10S1779 and D10S226 from the 10p15 region (Fig. 1a), revealed an absence of an allele at D10S1779 in the affected individuals (II-2, II-4, III-3 and III-4) (Fig. 1c). FISH analysis using PACs and BACs from the contig (Fig. 1a) spanning the D10S1779 locus confirmed the presence of a microdeletion in the four affected family members (data not shown). The size of the deletion was estimated as 250 kb (Fig. 1a), with the telomeric and centromeric breakpoints being within BAC457F3 and PACdJ1119O21, respectively. This deletion overlaps with that of patient 5/99, and the combined results from patient 5/99 and family 23/91 indicated that the critical HDR deletion region is about 200 kb in size (Fig. 1a).

A search for candidate genes revealed that *GATA3*, which has an expression pattern that includes the human embryonic kidney¹¹, parathyroids and inner ear¹², is located on 10p14-15 (ref. 13). The use of specific primers revealed that *GATA3* is contained in BAC49M24, YAC927G5 and YAC781B12, locating it within the critical interval. In addition, FISH analysis using cosmid clone 1.2 (ref. 7), which contains the entire *GATA3* gene (Fig. 1a), showed that one *GATA3* allele is deleted in patient 5/99 (data not shown) and in the affected members from family 23/91 (Fig. 1d). This clearly supports the involvement of *GATA3* in causing HDR.

To establish the role of *GATA3*, we carried out mutational analysis in five unrelated HDR patients who lacked cytogenetic abnormalities. DNA sequence analysis of the *GATA3* gene revealed the presence of two intragenic deletions and one nonsense mutation (Table 1, Fig. 2) in three of the five HDR patients tested. Each of these mutations was confirmed by the use of appropriate PCR primers and the identification of a smaller PCR product, or by heteroduplex formation or restriction enzyme analysis. The 49-base pair (bp) intragenic deletion that results in a frameshift (Fig. 2a) and the nonsense mutation (Fig. 2b) are predicted to result in truncated proteins that lack both zinc fingers (Table 1, Fig. 3), thereby leading to *GATA3* haplo-insufficiency¹⁴. However, the consequences of the 12-bp in-frame deletion (Fig. 2c), which leads to a loss of four amino acids that includes the first cysteine of the carboxy-terminal

Table 1 *GATA3* abnormalities in HDR patients

Family/patient	Phenotype*	Location	Mutation†	Amino-acid change	Predicted effect
Gene deletions					
5/99	H, D, Ra	D10S1751-D10S1779	900-kb deletion	Deletion of one allele	Haplo-insufficiency
23/91 (4 patients)	H, D, Rd	D10S1779	250-kb deletion	Deletion of one allele	Haplo-insufficiency
Intragenic deletions					
26/99	H, D, Rd	Exon 3	465-513 deletion	Frameshift, from codon 156	Missense, 22 amino acids from 156-177 followed by a premature stop at codon 178. Truncated protein with loss of ZF1 and ZF2; HI
3/99	H, D, Rd	Exon 5	946-957 deletion	In-frame deletion of codons 316-319	Loss of four amino acids (TSCA) and disruption of ZF2; HI
Non-sense mutation					
12/99 (mother & son)	H, D,	Exon 4	828C→T	R277X	Truncated protein with loss of ZF1 and ZF2; HI

* All patients have primary hypoparathyroidism (H), bilateral sensorineural deafness (D) and renal anomalies (Ra, renal agenesis; Rd, renal dysplasia) with the exception of family 12/99 in which the mother and son (Fig. 2b) do not have renal anomalies. Family members were not available from patient 3/99. The clinical details from 5/99, 23/91 and 26/99 have been previously reported^{5,17}. All patients had a normal lymphocyte and T-cell count and no clinical history of recurrent infections. In patients 3/99, 5/99 and 26/99, T-cell differentiation (CD4⁺ and CD8⁺ counts) and immunoglobulin levels were also normal.

† Nucleotide numbers refer to *GATA3* cDNA sequence (Genbank accession number X55122). ZF, zinc finger; HI, haplo-insufficiency.

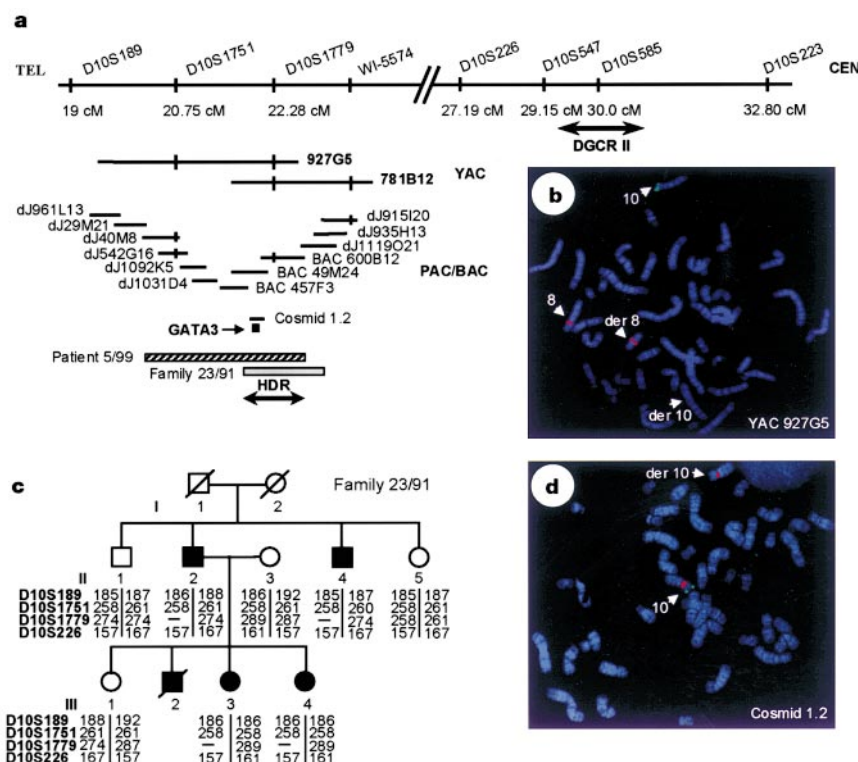


Figure 1 Physical and genetic map of the HDR region on 10p15. **a**, STSs and their genetic distances in centimorgans (cM) are indicated. The YAC/BAC/PAC contig that spans the two overlapping microdeletions in patient 5/99 (hatched bar) and family 23/91 (shaded bar) is shown. Critical regions for the HDR syndrome and DiGeorge critical region II (DGCR II) are indicated between the arrowheads. Cosmid 1.2 contains the *GATA3* gene and maps within the HDR critical region. **b**, FISH analysis using YAC927G5 on metaphase spreads of the translocation patient 5/99 showed only one normal signal (green dot), revealing a microdeletion. Red signals represent a chromosome-8-specific probe. **c**, Haplotype data

for markers D10S189, D10S1751, D10S1779 and D10S226 in family 23/91. Filled symbols represent affected individuals in whom a deleted allele for D10S1779 (–) was found. Individual III-1 has renal dysplasia without deafness or hypoparathyroidism³ and was therefore unlikely to have HDR, which is consistent with the presence of two alleles at D10S1779. **d**, FISH analysis using the *GATA3* cosmid 1.2 (green) on metaphase spreads of individual II-2 confirmed the deletion of *GATA3*. Red signals represent a chromosome-10-specific probe.

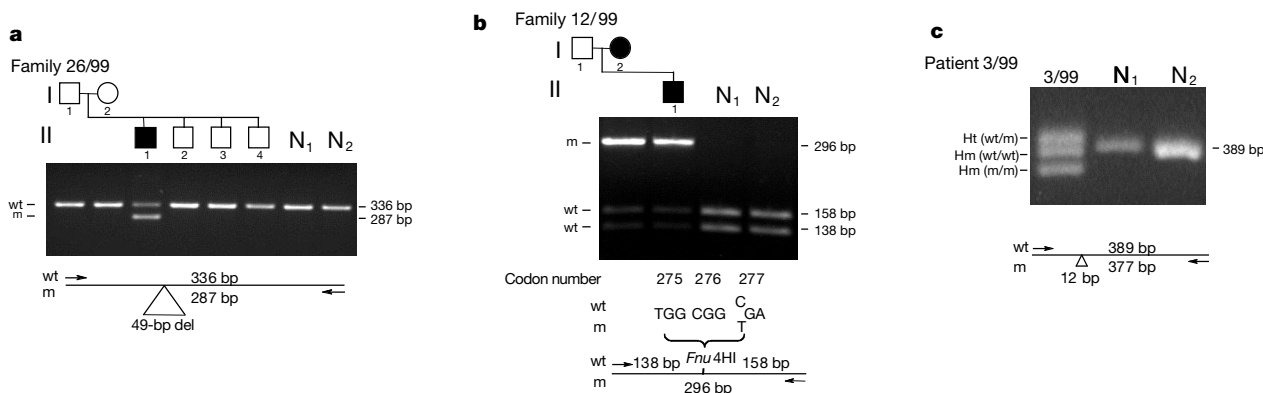


Figure 2 Detection of *GATA3* mutations in HDR patients. **a**, Deletional frameshift of 49 bp in family 26/99. **b**, Transition (C to T) resulting in a premature termination codon (Arg277→Stop) in family 12/99. **c**, In-frame deletion of 12 bp in patient 3/99. Top panels show co-segregation of the mutation and HDR; lower panels give sizes of the wild-type (wt) and mutant (m) PCR products. Mutations resulted in a smaller PCR product in family

26/99 (**a**), a loss of an *Fnu*4HI restriction site in family 12/99 (**b**) and the formation of homoduplexes (Hm, wt/wt or m/m), and heteroduplexes (Ht, wt/m) in patient 3/99 (**c**). The deletion in patient II-1 of family 26/99 (**a**) was shown to be a *de novo* mutation. These mutations were not present in 110 alleles from 55 unrelated normal individuals (N_1 and N_2 shown), which indicates that they are not common DNA sequence polymorphisms.

zinc finger (Fig. 3b), are more difficult to predict, and we therefore assessed this by DNA binding studies. Electrophoretic mobility shift assays (EMSAs), using nuclear extracts from cells transfected with either the wild-type or mutant *GATA3* (316–319del) construct (Fig. 3c), revealed that the mutation resulted in an absence of DNA binding, suggesting a loss of function. This is in agreement with human *GATA3* DNA-binding studies¹⁴ which have shown that the C-terminal zinc finger is essential for specific DNA binding. Thus, our results show that haplo-insufficiency of *GATA3* is the

underlying mechanism for the HDR syndrome. In addition, these findings emphasize the association between human developmental disorders and haplo-insufficiency of zinc-finger proteins. Other notable examples are holoprosencephaly¹⁵ and the tricho-rhino-phalangeal syndrome type I (ref. 16), which are caused by the *ZIC2* and *TRPS1* genes, respectively.

A detailed examination of the HDR phenotypes revealed that all patients with *GATA3* haplo-insufficiency have hypoparathyroidism and sensorineural deafness^{3,5,17}. The two patients with the nonsense

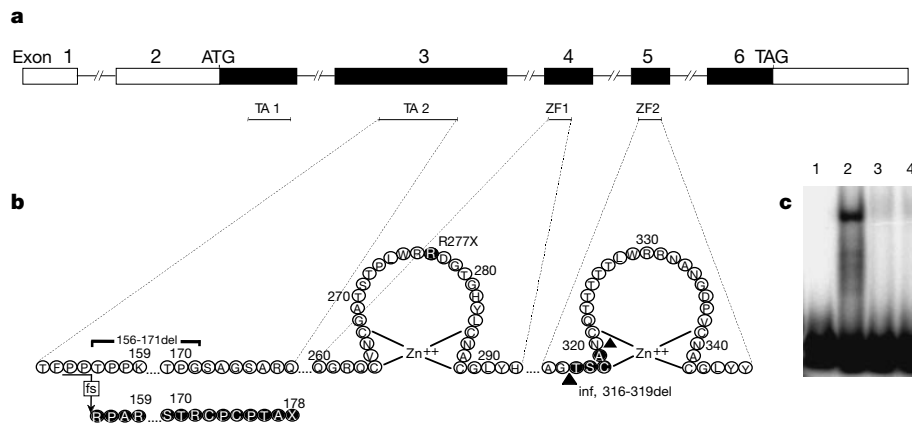


Figure 3 Locations of the *GATA3* mutations and effects on DNA binding. **a**, Genomic organization of the *GATA3* gene, which consists of six exons⁷ that encode two transactivating domains (TA1 and TA2) and two zinc fingers (ZF1 and ZF2)¹⁴. The ATG (start) site is in exon 2 and the TAG (stop) codon is in exon 6. **b**, Locations of the 49-bp deletional frameshift (fs) and the resulting missense peptide, together with the premature stop at codon 178 within TA2, the R277X mutation within ZF1 and the in-frame (inf) deletion (del) between arrowheads in ZF2. The altered amino acids are highlighted in

black, and every tenth amino acid is numbered. **c**, Analysis of the DNA-binding properties of the *GATA3* 316-319del mutant. Lane 1, control oligonucleotide without protein added; lane 2, binding of the wild-type *GATA3* protein; lane 3, absence of binding by mutant *GATA3*; lane 4, control nuclear extract from untransfected COS-1 cells. Expression of wild-type and mutant *GATA3* proteins was confirmed by western blot analysis (data not shown) of the nuclear extracts, using a polyclonal antiserum against *GATA3*.

mutation (Fig. 2b, Table 1) lack renal malformations, indicating a reduced penetrance for renal anomalies. The identification of additional patients with a partial HDR syndrome will help to delineate the clinical spectrum and possible genotype-phenotype relationships in this developmental disorder. We found no mutations in *GATA3* in two HDR patients examined, one of whom also has vaginal atresia as an atypical feature. These patients may have mutations in regulatory sequences flanking the *GATA3* gene or else they may represent genetic heterogeneity.

The HDR phenotype is consistent with the expression pattern of *GATA3* during human and mouse embryogenesis in the developing kidney (ureteric bud, collecting duct system and mesangial cells)¹¹, otic vesicle¹⁸ and parathyroids^{12,19}. But a more complex role for *GATA3* is suggested by its widespread expression in mice and man including the developing central nervous system and the haematopoietic organs^{12,19,20}. This is consistent with *GATA3* being essential in mouse T-cell development²¹ and with the total block of T-cell differentiation and several defects, especially of the central nervous system, observed in homozygous *GATA3* knockout mice⁸. None of our HDR patients with *GATA3* haplo-insufficiency had clinical or biochemical evidence of immune deficiency (Table 1)^{3,5,17}, indicating that the immune abnormalities observed in some patients with 10p deletions are most likely to be caused by other genes on 10p. Similarly, facial dysmorphism, growth and developmental delay, commonly seen in patients with a larger chromosome 10p deletion, were absent in the patients with *GATA3* mutations. This indicates that these features are also related to other genes on chromosome 10p. In contrast to humans, heterozygous *GATA3* knockout mice were not reported to have any anomalies associated with the HDR phenotype⁸. Thus, our finding that *GATA3* haplo-insufficiency causes HDR was not predicted from the mouse model, and a further re-examination of the heterozygous *GATA3* knockout mice for such abnormalities is required. This situation contrasts to that recently reported for *GATA1* abnormalities, which result in dyserythropoietic anaemia and thrombocytopenia²², in both mice and man. The mechanisms underlying these interspecies differences^{23,24} and the tissue-specific differences that confer susceptibility to *GATA3* haplo-insufficiency remain to be elucidated. Our finding that *GATA3* haplo-insufficiency causes a malformation syndrome shows the importance of this GATA factor in the formation of the inner ear, kidneys and parathyroids, and indicates that further studies of *GATA3* and its regulated genes will help to elucidate the developmental pathways in these organs. □

Methods

YAC, PAC, BAC and cosmid clones

The YAC contig was constructed and confirmed by combining Whitehead/MIT data (<http://www.genome.wi.mit.edu>), STS amplification, YAC end sequencing and *in situ* hybridization. YAC927G5 was used to screen the RPC11,3-5 Human PAC library. The PAC contig was generated using the Image 3.0 and FPC programs^{25,26} in collaboration with the Chromosome 10 Sequencing Group at the Sanger Centre (<http://www.sanger.ac.uk/HGP/Chr10>). PAC end sequencing allowed us to derive new STSs to screen the Human BAC library (Research Genetics) to fill the gaps in our PAC contig.

Haplotype analysis

Primer pairs for microsatellite (CA)_n markers were obtained from Whitehead/MIT and used to amplify 50 ng of genomic DNA at recommended conditions. PCR products were analysed using Allele Links (Amersham Pharmacia Biotech) and Cyrillic 2.13 (Cherwell Scientific).

Mutational analysis

Nine pairs of *GATA3*-specific primers (<http://www.kuleuven.ac.be/med/molonc/hdr.htm>) were used to amplify the six exons and intron-exon boundaries (Fig. 3a) from genomic DNA (250 ng). The DNA sequences of both strands were determined by *Taq* polymerase cycle sequencing²⁷.

EMSA

COS-1 cells were transfected with either the *GATA3* wild-type or the 316-319del mutant construct. Nuclear extracts from these transfected cells were prepared²⁸ and binding reactions were performed using a double-stranded oligonucleotide containing the *GATA3* consensus as described¹⁴. Western blot analysis was performed using polyclonal antiserum against *GATA3* (Santa Cruz).

Received 4 February; accepted 23 May 2000.

1. Daw, S. C. *et al.* A common region of 10p deleted in DiGeorge and velocardiofacial syndromes. *Nature Genet.* **13**, 458–460 (1996).
2. Schuffenhauer, S. *et al.* Deletion mapping on chromosome 10p and definition of a critical region for the second DiGeorge syndrome locus (DGS2). *Eur. J. Hum. Genet.* **6**, 213–225 (1998).
3. Bilous, R. W. *et al.* Autosomal dominant familial hypoparathyroidism, sensorineural deafness, and renal dysplasia. *N. Engl. J. Med.* **327**, 1069–1074 (1992).
4. McKusick, V. A. *Mendelian Inheritance in Man. Catalogs of Human Genes and Genetic Disorders* (Johns Hopkins Univ. Press, Baltimore, 1998).
5. Van Esch, H. *et al.* Partial DiGeorge syndrome in two patients with a 10p rearrangement. *Clin. Genet.* **55**, 269–276 (1999).
6. Lichtner, P. *et al.* An HDR (hypoparathyroidism, deafness, renal dysplasia) syndrome locus maps distal to the DiGeorge syndrome region on 10p13/14. *J. Med. Genet.* **37**, 33–37 (2000).
7. Labastie, M. C. *et al.* Structure and expression of the human *GATA3* gene. *Genomics* **21**, 1–6 (1994).
8. Pandolfi, P. P. *et al.* Targeted disruption of the *GATA3* gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. *Nature Genet.* **11**, 40–44 (1995).
9. Kuo, C. T. *et al.* *GATA4* transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev.* **11**, 1048–1060 (1997).
10. Simon, M. C. Gotta have *GATA*. *Nature Genet.* **11**, 9–11 (1995).
11. Labastie, M. C., Catala, M., Gregoire, J. M. & Peault, B. The *GATA-3* gene is expressed during human kidney embryogenesis. *Kidney Int.* **47**, 1597–1603 (1995).

12. Debacker, C., Catala, M. & Labastie, M. C. Embryonic expression of the human GATA-3 gene. *Mech. Dev.* **85**, 183–187 (1999).
13. Schuler, G. D. *et al.* A gene map of the human genome. *Science* **274**, 540–546 (1996).
14. Yang, Z. *et al.* Human GATA-3 trans-activation, DNA-binding, and nuclear localization activities are organized into distinct structural domains. *Mol. Cell. Biol.* **14**, 2201–2212 (1994).
15. Brown, S. A. *et al.* Holoprosencephaly due to mutations in ZIC2, a homologue of *Drosophila* odd-paired. *Nature Genet.* **20**, 180–183 (1998).
16. Momeni, P. *et al.* Mutations in a new gene, encoding a zinc-finger protein, cause tricho-rhino-phalangeal syndrome type I. *Nature Genet.* **24**, 71–74 (2000).
17. Beetz, R. *et al.* Hypoparathyroidism und Innenohrschwerhörigkeit. *Monatsschrift Kinderheilkunde* **145**, 347–352 (1997).
18. Rivolta, M. N. & Holley, M. C. GATA3 is downregulated during hair cell differentiation in the mouse cochlea. *J. Neurocytol.* **27**, 637–647 (1998).
19. George, K. M. *et al.* Embryonic expression and cloning of the murine GATA-3 gene. *Development* **120**, 2673–2686 (1994).
20. Hendriks, R. W. *et al.* Expression of the transcription factor GATA-3 is required for the development of the earliest T cell progenitors and correlates with stages of cellular proliferation in the thymus. *Eur. J. Immunol.* **29**, 1912–1918 (1999).
21. Ting, C. N., Olson, M. C., Barton, K. P. & Leiden, J. M. Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature* **384**, 474–478 (1996).
22. Nichols, K. E. *et al.* Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nature Genet.* **24**, 266–270 (2000).
23. Nanni, L. *et al.* The mutational spectrum of the sonic hedgehog gene in holoprosencephaly: SHH mutations cause a significant proportion of autosomal dominant holoprosencephaly. *Hum. Mol. Genet.* **8**, 2479–2488 (1999).
24. Rosenthal, N. & Harvey, R. P. Single allele mutations at the heart of congenital disease. *J. Clin. Invest.* **104**, 1483–1484 (1999).
25. Soderlund, C., Longden, I. & Mott, R. FPC: a system for building contigs from restriction fingerprinted clones. *Comput. Appl. Biosci.* **13**, 523–535 (1997).
26. Sulston, J., Mallett, F., Durbin, R. & Horsnell, T. Image analysis of restriction enzyme fingerprint autoradiograms. *Comput. Appl. Biosci.* **5**, 101–106 (1989).
27. Lloyd, S. E. *et al.* A common molecular basis for three inherited kidney stone diseases. *Nature* **379**, 445–449 (1996).
28. Latchman, D. S. (ed.) *Transcription Factors* (Oxford Univ. Press, Oxford, 1993).

Acknowledgements

We are grateful to R. Thoenen for the FISH analysis; A. Poffyn for clinical data; P. Romeo for the gift of cosmid clone 1.2; T. Meitinger for his support; the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO), the “Geconcerteerde Onderzoeksactie 1997–2001” and the Interuniversitaire Attractie Polen (IUAP) for support; the Medical Research Council, UK, for support (M.A.N., B.H., R.V.T.). H.V.E. is an Aspirant and K.D. is a Senior Clinical Investigator of the Fund for Scientific Research–Flanders, Belgium (FWO–Vlaanderen).

Correspondence and requests for materials should be addressed to K.D.
(e-mail: koen.devriendt@med.kuleuven.ac.be) or R.V.T.
(e-mail: rajesh.thakker@ndm.ox.ac.uk).

A motif in the $\alpha\beta$ T-cell receptor controls positive selection by modulating ERK activity

Guy Werlen, Barbara Hausmann & Ed Palmer

Basel Institute for Immunology, Grenzacherstrasse 487, CH-4005 Basel, Switzerland

Positive selection allows thymocytes that recognize an individual's own major histocompatibility complex (self-MHC) molecules to survive and differentiate, whereas negative selection removes overtly self-reactive thymocytes¹. Although both forms of thymic selection are mediated by the $\alpha\beta$ T-cell receptor (TCR) and require self-MHC recognition, an important question is whether they are controlled by distinct signalling cascades². We have shown that mutation of an essential motif within the TCR α -chain-connecting peptide domain (α -CPM) profoundly affects positive but not negative selection³. Using transgenic mice expressing a mutant α -CPM TCR we examined the contribution of several mitogen-activated protein kinase (MAPK) cascades to thymic selection. Here we show that in thymocytes expressing a mutant α -CPM receptor, a positively selecting peptide failed to

activate the extracellular signal-regulated kinase (ERK), although other MAPK cascades were induced normally. The defect in ERK activation was associated with impaired recruitment of the activated tyrosine kinases Lck and ZAP-70, phosphorylated forms of the TCR component CD3 ζ and the adaptor protein LAT to detergent-insoluble glycolipid-enriched microdomains (DIGs). Therefore, an intact DIG-associated signalosome is essential for sustained ERK activation, which leads to positive selection.

To investigate the role of the α -CPM in mediating positive selection, we studied Rag-2^{-/-} mice expressing transgenic wild-type or α -CPM mutant OT-1 TCRs. These receptors are referred to as wild-type and mutant, respectively. The OT-1 TCR specifically recognizes SIINFEKL, an octameric peptide from ovalbumin (OVA_{257–264}) presented by the class I MHC molecule, H-2K^b (ref. 4). The wild-type OT-1 thymus contains significant numbers of CD4⁺CD8⁺ single-positive (SP) thymocytes, as expected for a class I MHC-restricted TCR (Fig. 1a and ref. 4). In contrast, thymi from mice expressing the mutant receptor contain CD4⁺CD8⁺ double-negative (DN) and CD4⁺CD8⁺ double-positive (DP) thymocytes and very few SP cells. As observed with an α -CPM

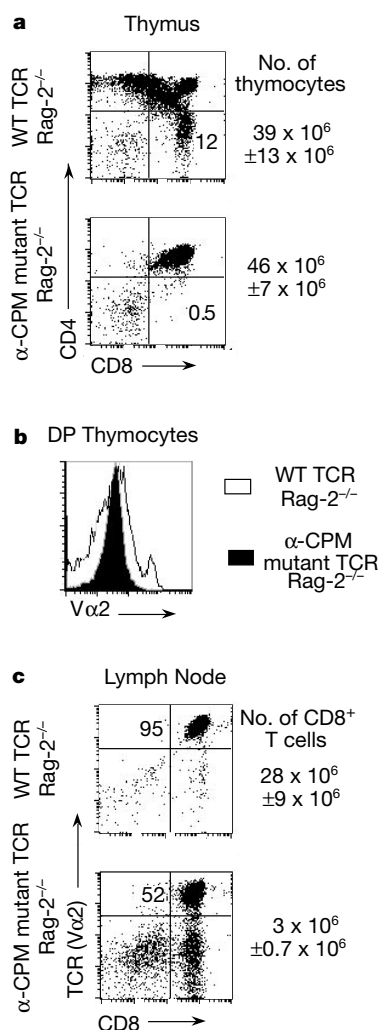


Figure 1 Analysis of thymocytes and peripheral cells from B6.Rag-2^{-/-} wild-type and mutant α -CPM mice. **a**, Thymocytes were stained with monoclonal antibodies against CD4 and CD8, and analysed by flow cytometry as described³. The numbers in the dot plots indicate the percentage of CD8⁺CD4⁺ SP thymocytes. WT, wild type. **b**, Surface expression of wild-type (white) and mutant (black) TCRs was assessed on DP thymocytes using flow cytometry after staining cells with an anti-V α 2 monoclonal antibody. **c**, Lymph node cells were stained with monoclonal antibodies against CD8 and V α 2. The numbers in the dot plots indicate the percentage of CD8⁺V α 2⁺ T cells among lymph node cells.

mutant class II MHC-restricted TCR³, the mutant OT-1 receptor is deficient in positive selection. A defect in TCR expression cannot account for the impaired positive selection, as DP thymocytes from wild-type and mutant mice bear comparable amounts of surface receptors (Fig. 1b). In the periphery, wild-type mice have a markedly higher percentage and number of CD8⁺ T cells than do mutant mice (Fig. 1c). As with mice expressing the α -CPM mutant class II MHC-restricted TCR³, the block in positive selection of DP thymocytes expressing the mutant α -CPM OT-1 receptor is leaky.

To study the ability of wild-type or mutant TCRs to transduce signals leading to positive or negative selection, we bred our transgenic mice to $\beta 2$ -microglobulin deficient mice ($\beta 2m^{-/-}$), which cannot generate CD4⁺CD8⁺ SP thymocytes owing to the absence of class I MHC molecules⁵. Analysis of thymi from $\beta 2m^{-/-}$ animals expressing the mutant or wild-type TCR showed developmental arrest at the DP stage (Fig. 2a). Most (> 90%) thymocytes from these mice express comparable levels of surface TCR (Fig. 2b). In fetal thymic organ cultures (FTOCs) derived from OT-1 transgenic mice, SIINFEKL induced negative selection whereas the related ligand, EIINFEKL, triggered positive selection^{4,6}. Using $\beta 2m^{-/-}$, Rag-2^{-/-} mice, we tested how the mutant thymocytes would respond to SIINFEKL or EIINFEKL by measuring the downregulation of CD4 and CD8 co-receptors in response to TCR stimulation⁷. Mutant and wild-type thymocytes responded similarly to SIINFEKL although mutant thymocytes are slightly less responsive (~14-fold) at low peptide concentrations (10 aM–1 pM) (Fig. 2c). Thus, cells expressing the mutant receptor can respond to an agonist, indicating that mutant α -CPM TCRs may be functional. However, mutant thymocytes were markedly less effective in responding to EIINFEKL, which required a 4,000-fold higher concentration to trigger an equivalent response to that observed in wild-type cells (Fig. 2d).

We also analysed thymocyte subpopulations in FTOCs derived from wild-type or mutant mice. The addition of EIINFEKL to wild-type thymic lobes resulted in a marked reduction in DP thymocytes and a substantial increase in CD4⁺CD8⁺ SP cells (Fig. 3), supporting previous studies⁴. In contrast, EIINFEKL did not significantly increase the number of SP thymocytes in mutant FTOCs, documenting the block of positive selection *in vitro* (Fig. 3). In the presence of SIINFEKL, wild-type and mutant FTOCs generated

similar proportions and numbers of DN and CD4⁺CD8⁺ SP thymocytes (Fig. 3). These SP cells are immature CD8⁺ cells that have not undergone positive selection⁶. Binding of the EIINFEKL/K^b and SIINFEKL/K^b tetramers to the mutant TCR was comparable to that seen with the wild-type receptor (data not shown), indicating that a deficiency in TCR–ligand interaction cannot account for the impaired positive selection. It seems more likely that the block in positive selection is due to impaired signalling generated on binding of EIINFEKL/K^b to the mutant TCR.

In thymocytes, the Ras/Raf/MEK/ERK cascade is thought to be involved in positive selection^{8–12}, whereas the MKK6/p38 pathway regulates negative selection¹³. The activation of the c-Jun-amino-terminal kinase (JNK) subfamily is relevant in peripheral T cells^{14–19}. To investigate whether a particular MAPK cascade is preferentially activated to promote positive or negative selection, we challenged wild-type or mutant thymocytes from Rag-2^{-/-}, $\beta 2m^{-/-}$ mice with SIINFEKL or EIINFEKL peptides, presented by glutaraldehyde-treated T2-K^b cells. Glutaraldehyde treatment eliminates MAPK activities from these antigen-presenting cells (APCs) (Figs 4c and 5c). Although 2 μ M SIINFEKL activated p38 in wild-type thymocytes (Fig. 4a), as expected for a peptide that promotes negative selection, SIINFEKL could also activate ERK and JNK. These results indicate that TCR engagement by an agonist does not discriminate between the MAPK signalling cascades. Interestingly, whereas the kinetics of JNK and p38 activation reached a plateau at 5–30 min, ERK activation by SIINFEKL was faster, peaking around 2 min and decreasing thereafter. SIINFEKL activated the MAPK cascades in mutant thymocytes with kinetics equivalent to those of wild-type cells (Fig. 4a). Comparing wild-type and mutant thymocytes, each MAPK activity showed a similar dose response to SIINFEKL stimulation (Fig. 4b). Together, these results indicate that the mutation of the α -CPM domain may not affect the activation of various MAPK cascades in response to SIINFEKL.

Does the α -CPM mutation affect the signal transduction induced by a ligand triggering positive selection? Although 2 μ M EIINFEKL induced sustained activation of ERK in wild-type thymocytes, no ERK activation was detected in mutant cells (Fig. 5a). Interestingly, EIINFEKL activated the p38 and JNK cascades similarly in mutant and wild-type thymocytes (Fig. 5a, b), indicating that mutation of the α -CPM domain specifically affects signalling to the ERK

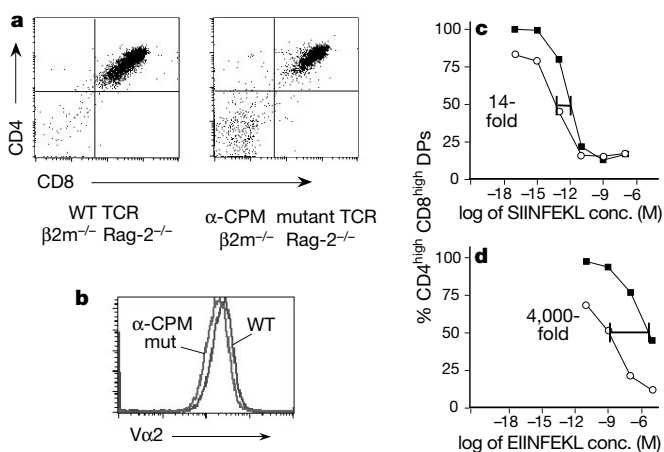


Figure 2 Analysis of CD4, CD8 and TCR expression in thymocytes from wild-type and mutant B6. $\beta 2m^{-/-}$, Rag-2^{-/-} mice. **a**, Thymocytes were stained with monoclonal antibodies and analysed as described in Fig. 1a. **b**, Surface expression of wild-type and mutant TCRs was assessed as described in Fig. 1b. **c**, **d**, Analysis of CD4/CD8 downregulation. Thymocytes bearing wild-type (open circles) or mutant TCRs (filled squares) were co-cultured for 16 h with T2-K^b antigen-presenting cells (APCs) in the presence of the indicated amounts of SIINFEKL (**c**) or EIINFEKL (**d**), respectively. Cells were collected, stained and analysed for CD4 and CD8 expression as described³.

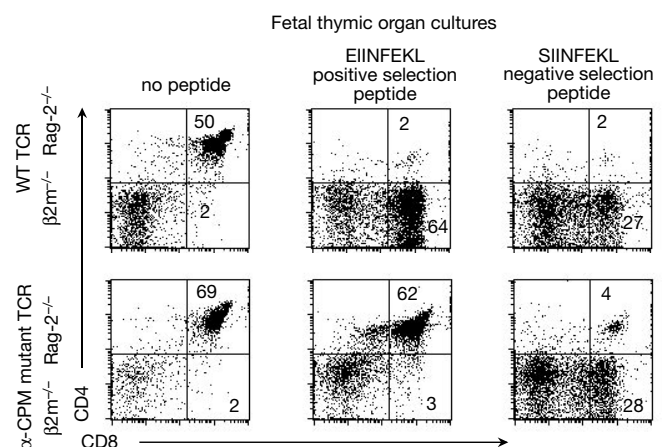


Figure 3 Analysis of thymocyte development induced by SIINFEKL or EIINFEKL in transgenic Rag-2^{-/-}, $\beta 2m^{-/-}$ fetal thymic organ cultures expressing wild-type or mutant TCRs. Thymic lobes were excised from fetal mice at gestational day 15 and cultured as described⁴ in the presence of 5 μ g ml⁻¹ human $\beta 2m$ and 20 μ M of EIINFEKL, SIINFEKL or no peptide. After 7 days of culture, thymocytes were collected, stained with monoclonal antibodies against V $\alpha 2$, CD4 and CD8, and analysed for CD4 and CD8 expression by flow cytometry. Numbers indicate the percentage of cells in the quadrant.

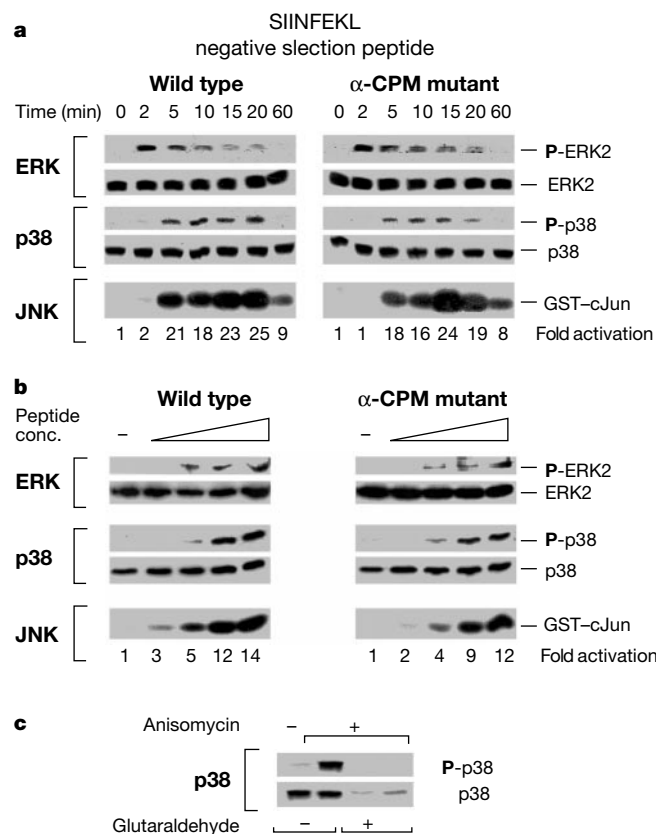


Figure 4 Analysis of MAPK activation induced by a negative selection peptide. **a**, Thymocytes were incubated for the indicated times with 2 μ M SIINFELK presented by glutaraldehyde-treated T2-K^b cells. Induction of p38 and ERK phosphorylation was analysed by immunoblotting. JNK activity was measured using an immunocomplex kinase assay¹⁷. **b**, Thymocytes were stimulated with 0, 0.2 pM, 200 pM, 20 nM or 2 μ M SIINFELK. Induction of ERK activity was analysed after 2 min; p38 and JNK activities were measured after 15 min of stimulation. **c**, Glutaraldehyde treatment eliminates MAPK activities from the APCs. Control and glutaraldehyde-treated T2-K^b cells were stimulated for 30 min with 2 μ M anisomycin, and p38 phosphorylation was analysed by immunoblotting.

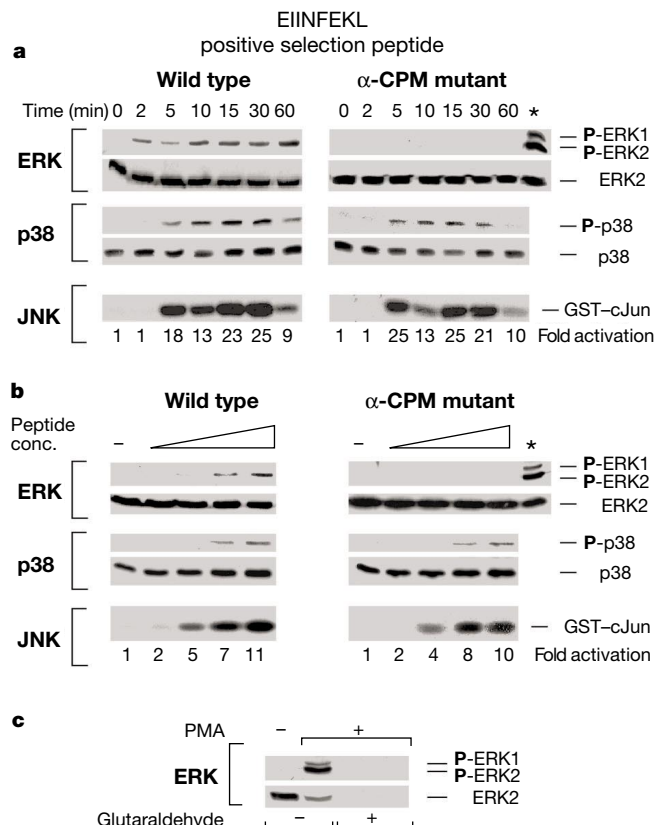


Figure 5 Analysis of MAPK activation induced by a positive selection peptide. **a**, Thymocytes were stimulated with 2 μ M EIINFELK for the indicated times. Activation of ERK, p38 or JNK was measured as described in Fig. 4a. Stimulation of thymocytes with 100 nM PMA was used as a positive control for ERK activation (asterisk). **b**, Thymocytes were stimulated with 0, 200 pM, 20 nM, 200 nM or 2 μ M EIINFELK. Activation of ERK, p38 or JNK was analysed as described in Fig. 4b. **c**, Glutaraldehyde treatment eliminates MAPK activities from the APCs. Control or glutaraldehyde-treated T2-K^b cells were stimulated for 15 min with 100 nM PMA, and ERK phosphorylation was assessed by immunoblotting.

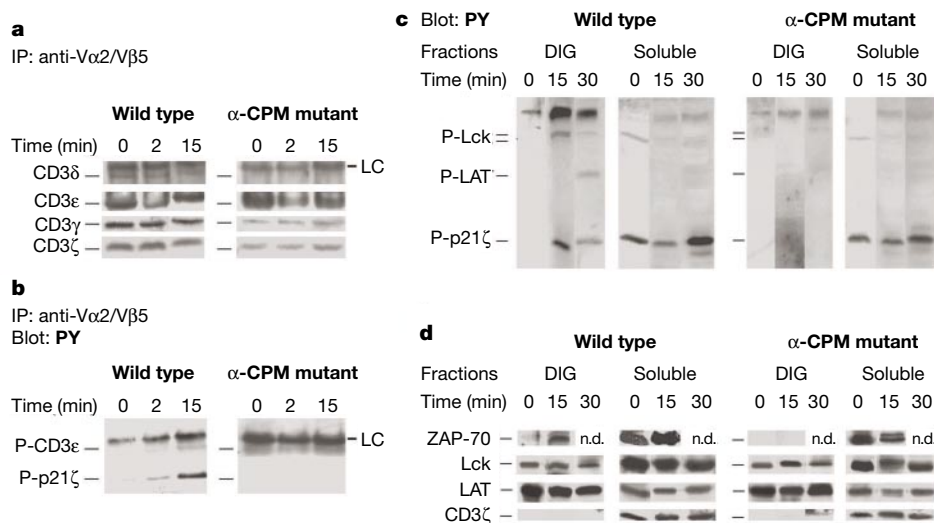


Figure 6 Analysis of proximal signalling events. Thymocytes (5×10^7) were stimulated for the indicated times with 2 μ M EIINFELK. The composition (**a**) and tyrosine-phosphorylation (**b**) of the immunoprecipitated TCR complex were analysed by immunoblotting as described³. The position of the Ig light chain (LC) from the antibody used for immunoprecipitation is indicated. **c**, Thymocytes were stimulated with 2 μ M

EIINFELK. The Brij58-containing cell lysates were fractionated as described²³; proteins were concentrated, resolved by SDS-PAGE and analysed by immunoblotting with an anti-phosphotyrosine antibody. **d**, Membranes were stripped and separately reprobed with anti-ZAP-70 monoclonal antibody, anti-Lck, anti-CD3 ζ ³ and anti-LAT²³ antibodies. n.d., not determined.

pathway. In wild-type thymocytes both the positive- and negative-selecting peptides activate the ERK pathway, but with different kinetics (Figs 4a and 5a). Thus, sustained ERK activation may be a prerequisite for positive selection. As p38 and JNK are equivalently activated by both peptides in wild-type and mutant thymocytes, they do not appear to be important for positive selection (Figs 4, 5), in agreement with previous work^{13–16,19}.

What components link the $\alpha\beta$ TCR to sustained activation of the ERK pathway? We found that all of the CD3 subunits were expressed in mutant thymocytes (data not shown). However, as we previously observed³, the mutant TCR failed to co-immunoprecipitate CD3 δ (Fig. 6a), a CD3 subunit implicated in the generation of mature T lymphocytes²⁰. Unlike the wild-type TCR, the mutant receptor was unable to co-immunoprecipitate P-p21 ζ (Fig. 6b), even though this active form of CD3 ζ is found in mutant thymocytes (Fig. 6c, soluble fraction). Propagation of TCR-derived signals requires the formation of a DIG-associated complex containing signalling and adaptor proteins^{21–23}. Analysis of tyrosine-phosphorylated proteins revealed that in wild-type thymocytes, a phosphorylated form of Lck (P-Lck), P-p21 ζ and a broad band with relative molecular mass $\sim 75,000$ (probably comprising ZAP-70 and SLP-76) were transiently recruited to the DIGs after 15 min of EIINFEKL stimulation (Fig. 6c). Although LAT, an adaptor protein required for complete activation of ERK²⁴, was constitutively associated with the DIGs (Fig. 6d), phospho-LAT appeared in lipid rafts only after 30 min of peptide stimulation (Fig. 6c). This indicates that LAT phosphorylation by ZAP-70 and Lck may ensue after these tyrosine kinases appear in the DIG-localized signalosome. In contrast, in mutant thymocytes, neither activated Lck, P-p21 ζ nor ZAP-70 are recruited to DIGs, which may explain why LAT is not phosphorylated despite its constitutive association with DIGs (Fig. 6c, d). The signalosome formed subsequent to EIINFEKL stimulation of mutant thymocytes is incomplete and probably accounts for the defective ERK activation. Together, our observations support a model in which a low-avidity ligand uses the α -CPM and CD3 δ to promote recruitment of activated Lck, P-p21 ζ and ZAP-70 to a DIG-associated signalosome that includes LAT. Phosphorylation of LAT may allow the recruitment to the lipid raft of components specifically activating ERK for a sustained period, which leads to positive selection. The activation of the JNK and p38 pathways that potentially mediate negative selection are not affected by the α -CPM mutation.

There is increasing evidence that cell survival or death in a number of cellular systems depends on the strength, duration and balance of various MAPK cascades^{25–27}. In PC12 pheochromocytoma cells, transient activation of ERK induced by epidermal growth factor stimulates proliferation whereas sustained ERK activation triggered by nerve growth factor induces differentiation^{25,26}. The MAPKs target different substrates, which might explain the selectivity of the ERK pathway in positive selection. Our results provide evidence that the TCR, by distinguishing the affinity of the ligand, modulates the composition of a DIG-associated signalling complex that controls the triggering of the ERK pathway, whereby sustained activation leads to positive selection and thymocyte survival. □

Methods

DNA constructs and transgenic mice

The sequence of the V α 2 and V β 5 chain complementary DNAs encoding the wild-type OT-1 receptor have been described⁴. The mutant α -CPM OT-1 TCR comprises a chimaeric α IV chain that completely lacks the α -CPM and a chimaeric β III chain as described (Fig. 1 in ref. 28). In the α IV construct the amino acids carboxy-terminal to the interchain Cys are derived from C δ , whereas in the β III construct the 28 C-terminal amino acids of the β -chain were replaced with homologous sequences from C γ 1 (Fig. 1 in ref. 28). cDNAs were cloned into the pHSE3' expression vector and the relevant DNA fragments were injected into C57BL/6 zygotes as described³.

Immunostaining and reagents

Freshly collected thymocytes from 7-week-old wild-type and mutant B6.Rag-2^{-/-} or wild-

type and mutant B6. β 2m^{-/-}, Rag-2^{-/-} transgenic mice were stained with monoclonal antibodies (mAbs) against V α 2, CD4 and CD8 and analysed by flow cytometry on a FACScan using the CellQuest software (Becton Dickinson) as described³. Anti-CD4, anti-CD8, anti-V β 5 and anti-V α 2 mAbs and anti-Lck antiserum were purchased from PharMingen and anti-ZAP-70 mAb from Transduction Laboratories. The anti-CD3 ζ mAb was used as described³. SIINFEKL and EIINFEKL peptides were synthesized as described²⁹.

Fetal thymic organ cultures (FTOCs)

FTOCs were carried out and analysed as described⁴.

Thymocyte stimulation and kinase activity assays

Antigen-presenting cells, T2-K^b (ref. 30) (2×10^6 ml⁻¹) were loaded in IMDM/5% FCS for 2 h at 37 °C with the indicated amounts of SIINFEKL or EIINFEKL, washed twice in PBS and fixed for 1 min in 0.1% glutaraldehyde in PBS. The reaction was stopped by adding 1 volume PBS containing 0.2 M lysine, and the cells were subsequently washed three times in PBS. Glutaraldehyde treatment of T2-K^b cells eliminates MAPK activation from these cells even when they are stimulated by potent pharmacological agonists (Figs 4c and 5c; and data not shown). After 4–5 h pre-incubation at 37 °C, 5×10^6 thymocytes were mixed with 5×10^6 peptide-loaded, glutaraldehyde-treated and pre-warmed T2-K^b cells and stimulated for the indicated times. Cells were lysed in whole-cell extract lysis buffer¹⁸ and kinase activities were analysed by immunoblotting using antiphospho-p38 antisera or the antiphospho-ERK1/2 mAb, E10 (NEB). JNK activity was measured by immunocomplex kinase assay using mAb G151-333 (PharMingen) to immunoprecipitate JNK1 and GST-cJun 1–79 as a substrate¹⁷. The phosphorylation of GST-cJun was quantified using a phosphorimager (Molecular Dynamics). Total levels of p38 or ERK expression were measured using anti-p38 or anti-ERK2 antisera (Santa-Cruz).

Immunoprecipitation and subcellular fractionation

Thymocytes were stimulated as described above, the TCR/CD3 complex was immunoprecipitated with anti-V α 2 (B20.1) and anti-V β 5 (MR9-4) from thymocyte lysates, and the CD3 isoforms were analysed by immunoblotting as described³. For analysis of DIG-associated signalosomes, thymocytes (2.5×10^8) were lysed on ice in 1 ml of 1% Brij58 lysis buffer²², homogenized 10 times and prepared for discontinuous sucrose gradient ultracentrifugation as described²³. After centrifugation, 0.4-ml fractions were collected from the top of the gradient. DIGs were recovered from low-density fractions 2–4 and detergent-soluble proteins from fractions 8–11. Concentrated proteins²² were resolved by SDS-PAGE. Tyrosine-phosphorylation of proteins was determined by immunoblot using mAb 4G10 (Upstate Biotechnology).

Received 6 March; accepted 19 May 2000.

1. von Boehmer, H., Teh, H. S. & Kisielow, P. The thymus selects the useful, neglects the useless and destroys the harmful. *Immunol. Today* **10**, 57–61 (1989).
2. Mariathasan, S., Jones, R. G. & Ohashi, P. S. Signals involved in thymocyte positive and negative selection. *Semin. Immunol.* **11**, 263–272 (1999).
3. Bäckström, B. T., Müller, U., Hausmann, B. & Palmer, E. Positive selection through a motif in the $\alpha\beta$ T cell receptor. *Science* **281**, 835–838 (1998).
4. Hogquist, K. A. *et al.* T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17–27 (1994).
5. Zijlstra, M. *et al.* β 2-microglobulin deficient mice lack CD4⁺8⁺ cytolytic T cells. *Nature* **344**, 742–746 (1990).
6. Hogquist, K. A., Jameson, S. C. & Bevan, M. J. Strong agonist ligands for the T cell receptor do not mediate positive selection of functional CD8⁺ T cells. *Immunity* **3**, 79–86 (1995).
7. Swat, W., Ignatowicz, L., von Boehmer, H. & Kisielow, P. Clonal deletion of immature CD4⁺8⁺ thymocytes in suspension culture by extrathymic antigen-presenting cells. *Nature* **351**, 150–153 (1991).
8. Alberola-Ila, J., Forbush, K. A., Seger, R., Krebs, E. G. & Perlmutter, R. M. Selective requirement for MAP kinase activation in thymocyte differentiation. *Nature* **373**, 620–623 (1995).
9. Alberola-Ila, J., Hogquist, K. A., Swan, K. A., Bevan, M. J. & Perlmutter, R. M. Positive and negative selection invoke distinct signaling pathways. *J. Exp. Med.* **184**, 9–18 (1996).
10. Swan, K. A. *et al.* Involvement of p21ras distinguishes positive and negative selection in thymocytes. *EMBO J.* **14**, 276–285 (1995).
11. O'Shea, C. C., Crompton, T., Rosewell, I. R., Hayday, A. C. & Owen, M. J. Raf regulates positive selection. *Eur. J. Immunol.* **26**, 2350–2355 (1996).
12. Pagès, G. *et al.* Defective thymocyte maturation in p44 MAP kinase (Erk 1) knockout mice. *Science* **286**, 1374–1377 (1999).
13. Sugawara, T., Moriguchi, T., Nishida, E. & Takahama, Y. Differential roles of ERK and p38 MAP kinase pathways in positive and negative selection of T lymphocytes. *Immunity* **9**, 565–574 (1998).
14. Yang, D. D. *et al.* Differentiation of CD4⁺ T cells to Th1 cells requires MAP kinase JNK2. *Immunity* **9**, 575–585 (1998).
15. Sabapathy, K. *et al.* JNK2 is required for efficient T-cell activation and apoptosis but not for normal lymphocyte development. *Curr. Biol.* **9**, 116–125 (1999).
16. Dong, C. *et al.* Defective T cell differentiation in the absence of Jnk1. *Science* **282**, 2092–2095 (1998).
17. Jacinto, E., Werlen, G. & Karin, M. Cooperation between Syk and Rac1 leads to synergistic JNK activation in T lymphocytes. *Immunity* **8**, 31–41 (1998).
18. Werlen, G., Jacinto, E., Xia, Y. & Karin, M. Calcineurin preferentially synergizes with PKC- θ to activate JNK and IL-2 promoter in T lymphocytes. *EMBO J.* **17**, 3101–3111 (1998).
19. Dong, C. *et al.* JNK is required for effector T-cell function but not for T-cell activation. *Nature* **405**, 91–94 (2000).
20. Dave, V. P. *et al.* CD3 δ deficiency arrests development of the $\alpha\beta$ but not the $\gamma\delta$ T cell lineage. *EMBO J.* **16**, 1360–1370 (1997).
21. Simons, K. & Ikonen, E. Functional rafts in cell membranes. *Nature* **387**, 569–572 (1997).
22. Montixi, C. *et al.* Engagement of T cell receptor triggers its recruitment to low-density detergent-insoluble membrane domains. *EMBO J.* **17**, 5334–5348 (1998).

23. Zhang, W., Tribble, R. P. & Samelson, L. E. LAT palmitoylation: its essential role in membrane microdomain targeting and tyrosine phosphorylation during T cell activation. *Immunity* **9**, 239–246 (1998).
24. Finco, T. S., Kadlecik, T., Zhang, W., Samelson, L. E. & Weiss, A. LAT is required for TCR-mediated activation of PLC γ and the Ras pathway. *Immunity* **9**, 617–626 (1998).
25. Marshall, C. J. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80**, 179–185 (1995).
26. York, R. D. *et al.* Rap1 mediates sustained MAP kinase activation induced by nerve growth factor. *Nature* **392**, 622–626 (1998).
27. Xia, Z., Dickens, M., Raingeaud, J., Davis, R. J. & Greenberg, M. E. Opposing effects of ERK and JNK-p38 MAP kinases on apoptosis. *Science* **270**, 1326–1331 (1995).
28. Bäckström, B. T. *et al.* A motif within the T cell receptor α chain constant region connecting peptide domain controls antigen responsiveness. *Immunity* **5**, 437–447 (1996).
29. Stotz, S. H., Bolliger, L., Carbone, F. R. & Palmer, E. T cell receptor (TCR) antagonism without a negative signal: evidence from T cell hybridomas expressing two independent TCRs. *J. Exp. Med.* **189**, 253–264 (1999).
30. Alexander, J., Payne, J. A., Murray, R., Frelinger, J. A. & Cresswell, P. Differential transport requirements of HLA and H-2 class I glycoproteins. *Immunogenetics* **29**, 380–388 (1989).

Acknowledgements

We thank F. Carbone, M. J. Bevan and D. Kioussis for transgenic mice expressing wild-type OT-1 receptors; A. Peter for generating mutant α -CPM TCR constructs; U. Müller for generation of mutant transgenic mice; T. Potter for T2-K^b cells; M. Daniels, K. Hogquist and S. Jameson for tetramers; L. Samelson for LAT antisera; S. Stotz and T. Harder for discussions; and E. Jacinto, P. Baldari, P. Kiselow and J. Kirberg for reading the manuscript. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-LaRoche Ltd.

Correspondence and requests for materials should be addressed to G.W. (e-mail: werlen@bii.ch) or E.P. (e-mail: palmer@bii.ch).

CD3 δ couples T-cell receptor signalling to ERK activation and thymocyte positive selection

Pilar Delgado[†], Edgar Fernández[†], Vibhuti Dave[‡], Dietmar Kappes[‡] & Balbino Alarcón^{*}

^{*} Centro de Biología Molecular Severo Ochoa, Consejo Superior de Investigaciones Científicas, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

[‡] Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, Pennsylvania 19111, USA

[†] These authors contributed equally to this work

Thymocytes from mice lacking the CD3 δ chain of the T-cell receptor (TCR), unlike those of other CD3-deficient mice^{1,2}, progress from a CD4⁺CD8[−] double-negative to a CD4⁺CD8⁺ double-positive stage. However, CD3 δ ^{−/−} double-positive cells fail to undergo positive selection, by which double-positive cells differentiate into more mature thymocytes³. Positive selection is also impaired in mice expressing inactive components of the Ras/mitogen activated protein (MAP) kinase signalling pathway^{4–6}. Here we show that CD3 δ ^{−/−} thymocytes are defective in the induction of extracellular signal-regulated protein kinase (ERK) MAP kinases upon TCR engagement, whereas activation of other MAP kinases is unaffected. The requirement for CD3 δ maps to its extracellular or transmembrane domains, or both, as expression of a tail-less CD3 δ rescues both ERK activation and positive selection in CD3 δ ^{−/−} mice. Furthermore, the defect correlates with severely impaired tyrosine phosphorylation of the linker protein LAT, and of the CD3 ζ chain that is localized to membrane lipid rafts upon TCR engagement. Our data indicate that the blockade of positive selection of CD3 δ ^{−/−} thymocytes may derive from defective tyrosine phosphorylation of CD3 ζ in lipid rafts, resulting in impaired activation of the LAT/Ras/ERK pathway.

Thymocyte development in mice lacking CD3 δ is arrested at the double-positive stage with markedly diminished TCR expression³ (Fig. 1a). Whether this developmental blockade results from impaired TCR expression or from a specific signalling defect independent of surface expression of the TCR is unclear. To address

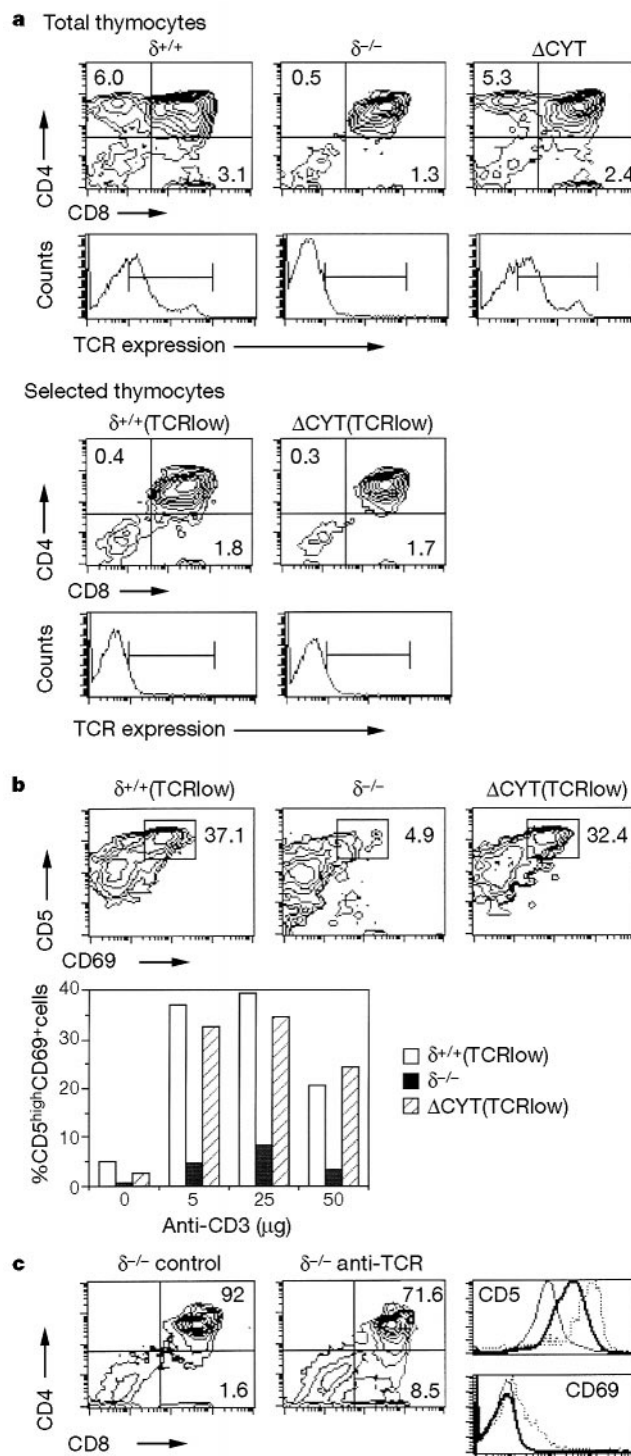


Figure 1 $\delta^{-/-}$ thymocytes are deficient in the induction of positive selection markers upon TCR engagement *in vitro* and *in vivo*. **a**, Flow cytometry analysis of CD4, CD8 and TCR expression in total or in TCR^{int/high}-depleted thymocytes (selected thymocytes). **b**, CD5 and CD69 expression on thymocytes treated with immobilized 145-2C11 antibody (5 μ g, top). Percentages of CD5^{high}CD69⁺ cells, gated as illustrated, are shown for treatment with the indicated amounts of 145-2C11 (bottom). **c**, Surface marker expression on total thymocytes from PBS-injected $\delta^{-/-}$ (control, thin solid line), anti-TCR-injected $\delta^{-/-}$ (thick solid line) or anti-TCR-injected $\delta^{+/+}$ (dotted line) mice.

this question, we selected immature thymocytes from $\delta^{+/+}$ and Δ CT mice ($\delta^{-/-}$ animals expressing a transgenic CD3 δ chain that lacks the whole cytoplasmic region⁷) with TCR expression comparable to that of $\delta^{-/-}$ thymocytes (Fig. 1a). Both TCR expression and positive selection are rescued in the Δ CT transgenic line⁷ (Fig. 1a). To investigate the functional competence of these different TCRs, $\delta^{-/-}$, $\delta^{+/+}$ (TCR^{low}) and Δ CT(TCR^{low}) thymocytes were treated *in vitro* with immobilized anti-TCR antibodies (Fig. 1b), or *in vivo* by injection of anti-TCR antibodies (Fig. 1c). TCR-mediated upregulation of CD5 and CD69 surface expression, events typically associated with positive selection⁸, was defective in $\delta^{-/-}$ compared with in both $\delta^{+/+}$ (TCR^{low}) and Δ CT(TCR^{low}) thymocytes (Fig. 1b, c), irrespective of the dose of antibody (Fig. 1b), and despite the induction of partial differentiation of double-positive thymocytes into CD8⁺ single-positive cells (Fig. 1c). These findings indicate that coupling to downstream signalling pathways may be altered in the absence of CD3 δ .

There is increasing evidence that Ras-mediated signalling pathways are critical for the regulation of CD69 expression in T cells⁹ and for positive selection of thymocytes^{4-6,10}. Specific suppression of ERK activity appears preferentially to block generation of CD4⁺ single-positive cells while favouring the appearance of CD8⁺ single-positive thymocytes^{11,12}. The similar bias observed in $\delta^{-/-}$ mice injected with anti-TCR antibodies (Fig. 1c) and the defect in TCR-mediated upregulation of CD69 (Fig. 1b, c) prompted us to investigate the capacity of CD3 δ -deficient TCRs to activate the MAP kinases ERK, c-Jun amino-terminal kinase (JNK) and p38 (ref. 13) in thymocytes. Analysis of the phosphorylation status of MAP kinases—an indication of their activation state¹³—upon TCR engagement revealed that ERK activation was markedly impaired in $\delta^{-/-}$ thymocytes, whereas induction of JNK and p38 was unaffected (Fig. 2a). Notably, Δ CT(TCR^{low}) thymocytes showed potent activation of ERK kinases following TCR engagement, comparable to that of $\delta^{+/+}$ (TCR^{low}) thymocytes (Fig. 2a). These results indicate that the absence of CD3 δ results in impaired activation of the ERK pathway unrelated to the level of TCR expression. Furthermore, the role of CD3 δ in ERK activation does not map to its cytoplasmic tail.

ERK is the last element in the MAP kinase cascade. It is initiated

when Raf is activated by the activated form of the small GTPase Ras¹³. We evaluated the levels of active (GTP-bound) Ras in $\delta^{-/-}$ and $\delta^{+/+}$ (TCR^{low}) thymocytes upon TCR engagement by using an immobilized glutathione S-transferase (GST) fusion protein containing the Ras-binding domain of Raf. The levels of Ras–GTP increased as early as 2 min after TCR engagement in $\delta^{+/+}$ (TCR^{low}) thymocytes (Fig. 2b) and were maintained for at least 15 min. In contrast, in spite of higher starting levels of Ras–GTP in resting $\delta^{-/-}$ thymocytes, TCR engagement resulted in a time-dependent decrease in Ras–GTP. These results indicate that in the absence of CD3 δ the Ras–ERK pathway may be altered at the initiation of the cascade.

Two leucocyte-specific proteins, the linker LAT and the adaptor SLP-76, serve as scaffolds for the integration of different signalling pathways¹⁴ and are essential for thymic development^{15,16}. SLP-76 was inducibly tyrosine phosphorylated in TCR-stimulated $\delta^{-/-}$ thymocytes to levels comparable to those of total $\delta^{+/+}$ thymocytes (Fig. 3a). In addition, induced tyrosine phosphorylation of Vav, an upstream regulator of the JNK and p38 MAP kinases¹⁷, was also comparable in $\delta^{+/+}$ (TCR^{low}) and $\delta^{-/-}$ thymocytes (Fig. 3b), and the level of phospho-Vav precipitating with SLP-76 was similar in stimulated $\delta^{+/+}$ and $\delta^{-/-}$ thymocytes (Fig. 3a). In contrast, induced tyrosine phosphorylation of SLP-76-associated LAT was severely diminished in $\delta^{-/-}$ thymocytes (Fig. 3a). These results indicate that $\delta^{-/-}$ thymocytes may be defective in the TCR-induced tyrosine phosphorylation of LAT but not in other protein tyrosine kinase (PTK) targets.

The levels of phospho-LAT in whole-cell lysates were lower in $\delta^{-/-}$ thymocytes than in $\delta^{+/+}$ (TCR^{low}) and Δ CT(TCR^{low}) thymocytes (Fig. 3c). Phospho-LAT, as detected by anti-phosphotyrosine antibody immunoprecipitation and anti-LAT immunoblotting, increased steadily following TCR engagement in $\delta^{+/+}$ (TCR^{low}) thymocytes. In contrast, the level of phospho-LAT was initially high in resting $\delta^{-/-}$ thymocytes but decreased after 2 min of activation (Fig. 3d). Reprobing the membrane with the anti-phosphotyrosine antibody revealed that there was less phospho-LAT in $\delta^{-/-}$ than in $\delta^{+/+}$ thymocytes even at time 0 of stimulation (Fig. 3d). This may indicate that in non-stimulated $\delta^{-/-}$ thymocytes non-phosphorylated LAT precipitates with other tyrosine-phosphorylated

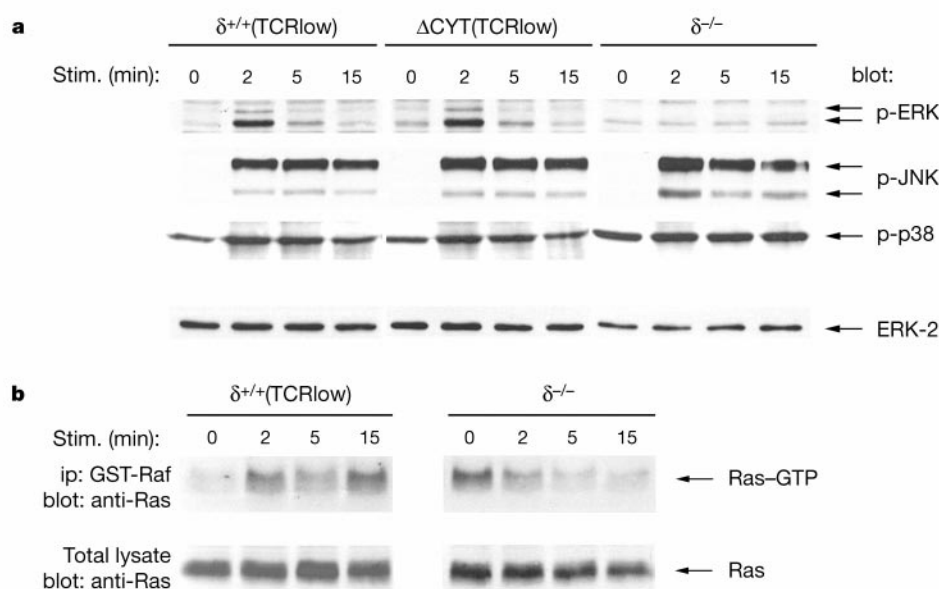


Figure 2 Impaired TCR-induced activation of ERK, but not JNK and p38 MAP kinases, in $\delta^{-/-}$ thymocytes. **a**, Following stimulation with crosslinked anti-CD3 (KT-3) and anti-CD4 (GK1.5) antibodies, we analysed induction of ERK, JNK and p38 phosphorylation by sequential immunoblotting of whole-cell lysates with phospho-MAP kinase-specific

antibodies. Protein loading was assessed by immunoblotting with a total ERK antibody. **b**, After stimulation with anti-CD3 antibody for the indicated times, we assessed the level of Ras–GTP by precipitation with Raf-1 RBD agarose followed by immunoblotting with anti-Ras antibody. Total Ras was measured by anti-Ras immunoblotting of total lysates.

proteins. Alternatively, it may indicate constitutive low-grade phosphorylation of LAT. Together, these results indicate that CD3 δ -deficient TCRs are not capable of signalling for adequate tyrosine phosphorylation of LAT. Furthermore, downstream events dependent on LAT tyrosine phosphorylation (phospholipase C γ -1 (PLC γ -1) phosphorylation and Ca²⁺ mobilization¹⁸) were impaired in $\delta^{-/-}$ compared with $\delta^{+/+}$ (TCR^{low}) thymocytes (Fig. 3e).

The defective TCR-induced CD69 upregulation and ERK activa-

tion shown here in CD3 δ -deficient thymocytes are similar to those reported in LAT-deficient T cells¹⁸. These findings support the proposal that CD3 δ somehow controls aspects of LAT function that are related to Ras/ERK pathway activation. Thus, it is puzzling that CD3 δ deficiency does not map to the immunoreceptor tyrosine-based activation motif (ITAM)-bearing cytoplasmic domain, but instead to the extracellular and/or transmembrane domains. This phenotype may be explained by a model in which the extra-

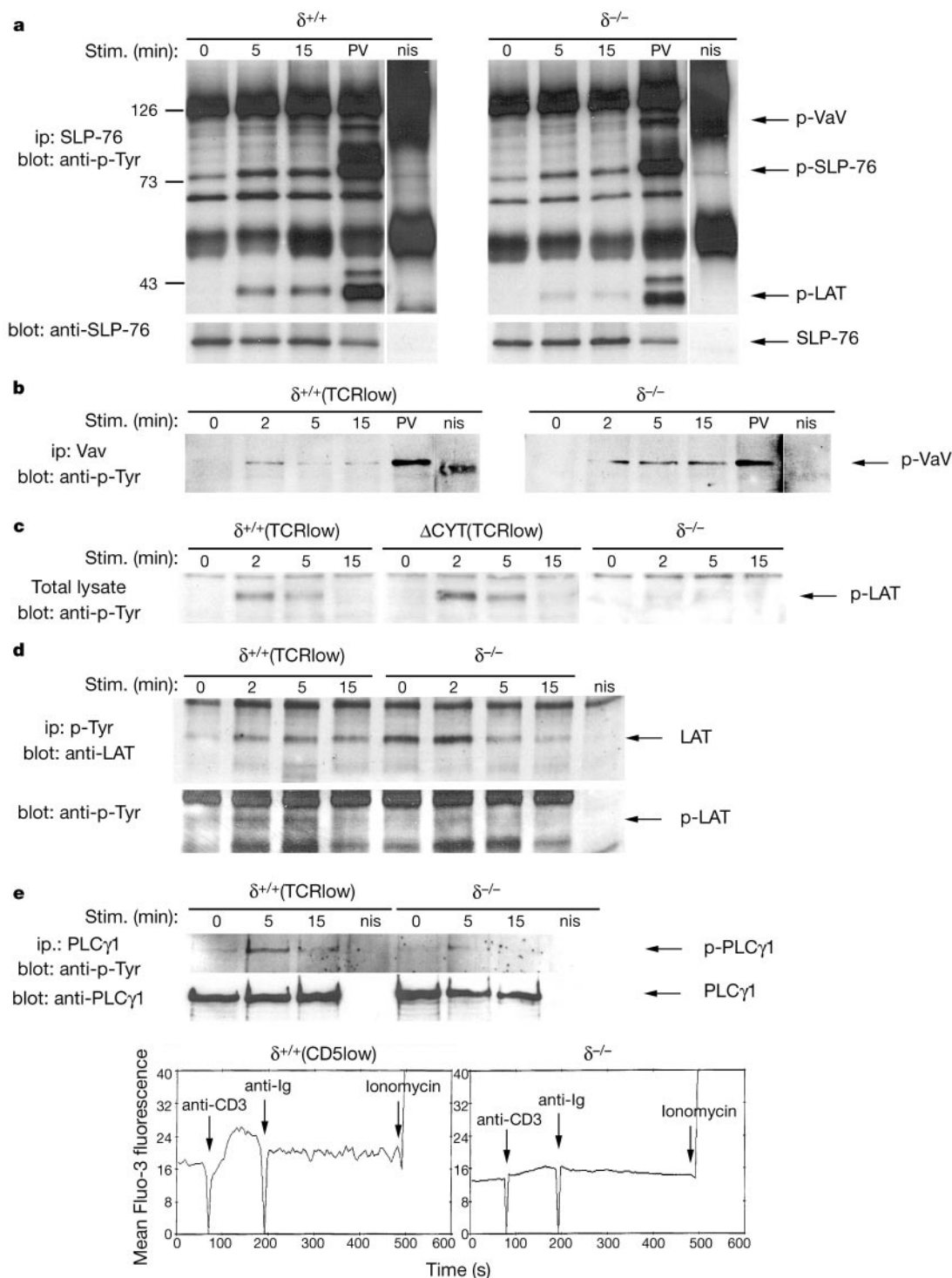


Figure 3 Impaired TCR-induced tyrosine phosphorylation of LAT, but not of SLP-76 or Vav, in $\delta^{-/-}$ thymocytes. Total (a) or selected (b–e) thymocytes were stimulated with either 145-2C11 for the times indicated or with pervanadate (PV) for 5 min. Cells were lysed in 1% Brij96 (a–c, e) or 1% Brij96 plus 60 mM octylglucoside (d). nis, non-immune

serum precipitates. Intracellular calcium fluxes (e) were measured in gated $\delta^{+/+}$ CD5^{low} (corresponding to TCR^{low} thymocytes²⁰) and $\delta^{-/-}$ thymocytes upon addition of anti-CD3 and crosslinking (anti-Ig) antibodies. Ionomycin (1 μ M) was added as a control for Fluo-3 AM loading.

cellular and transmembrane domains of CD3 δ are involved in specifically translocating the engaged TCR complex and its associated PTKs to lipid rafts^{19,20}, where LAT is constitutively localized²¹.

To test this hypothesis, we examined the distribution of TCR components in lipid rafts following receptor engagement in $\delta^{-/-}$ and $\delta^{+/+}$ (TCR^{low}) thymocytes. As shown in Fig. 4a, tyrosine-phosphorylated CD3 ϵ was present in lipid rafts (insoluble fraction) from both $\delta^{-/-}$ and $\delta^{+/+}$ (TCR^{low}) thymocytes. In contrast, compared with $\delta^{+/+}$ (TCR^{low}) thymocytes, tyrosine-phosphorylated CD3 ζ was markedly diminished in lipid rafts, but not in the soluble fraction, of TCR-stimulated $\delta^{-/-}$ thymocytes. The presence of the tail-less CD3 δ restored the levels of phospho-CD3 ζ in lipid rafts (Fig. 4b). The selective defect in tyrosine phosphorylation of lipid-raft-associated CD3 ζ in $\delta^{-/-}$ thymocytes did not result from abnormal distribution of the TCR, as the levels of total CD3 ζ in soluble and insoluble fractions resembled those in $\delta^{+/+}$ thymocytes (Fig. 4a, b). The uncoupling of TCR-initiated signals from appropriate tyrosine phosphorylation of CD3 ζ in lipid rafts could explain the signalling defects downstream of LAT phosphorylation selectively observed in $\delta^{-/-}$ thymocytes. Defective TCR-mediated tyrosine phosphorylation of CD3 ζ in lipid rafts of $\delta^{-/-}$ thymocytes may reflect a specific structural requirement of the extracellular and/or transmembrane domains of CD3 δ for effective multi-step tyrosine phosphorylation of CD3 ζ ²². Alternatively, CD3 δ may be required to protect phosphorylated CD3 ζ from the action of tyrosine phosphatases recruited to the engaged TCR²³.

The defective activation of the ERK pathway might explain the blockade of positive selection in mice lacking CD3 δ ³. Positive, but not negative, selection was also blocked in mice expressing a transgenic TCR mutated in a motif within the TCR α -connecting peptide domain²⁴. This mutation specifically disrupts the association of CD3 δ with the TCR complex, further supporting a role for CD3 δ in the regulation of signals involved in positive selection.

Given the presence of an ITAM motif in the cytoplasmic regions of the CD3 components of the TCR, redundant and specific roles for the CD3 chains are likely to coexist²⁵. Collectively, our data show that CD3 δ has a role—independent of its cytoplasmic domain—in selectively coupling the TCR to activation of the ERK signalling pathway, thus allowing positive selection of thymocytes. □

Methods

Materials

We used the following antibodies: sheep anti-SLP76 (gift from G. Koretzky); rabbit anti-LAT (gift from L. Samelson); hamster anti-murine CD3 monoclonal antibody 145-2C11 (gift from J. Bluestone); hamster anti-murine TCR β monoclonal antibody H57-597 (Pharmingen); rat anti-murine CD3 monoclonal antibody KT-3 (Serotec); rat anti-murine CD4 monoclonal antibody GK1.5 (Pharmingen); rabbit anti-rat and anti-hamster Ig antibodies (Jackson ImmunoResearch); cocktail of mouse anti-hamster monoclonal antibodies (Pharmingen); anti-Vav and anti-phospho-ERK antibody (Santa Cruz Biotechnology); anti-phospho-JNK, anti-phospho-p38 and total ERK (New England Biolabs); anti-phosphotyrosine monoclonal antibody 4G10 and anti-PLC- γ 1 (Upstate Biotechnology). The anti-CD3 ϵ monoclonal antibody APA1/1 and the rabbit anti-CD3 ζ antibody 448 were generated as described²⁶. All fluorescein-conjugated and phycoerythrin-conjugated antibodies were from Pharmingen.

Thymocyte stimulation and flow cytometry

For *in vitro* stimulation, we incubated 5×10^5 thymocytes for 24 h on culture wells pre-coated with various amounts of 145-2C11 or PBS as control. Cells were stained with specific fluorochrome-conjugated antibodies and analysed by flow cytometry in a FACSCalibur flow cytometer using the CellQuest software (Becton Dickinson). For *in vivo* stimulation, $\delta^{-/-}$ mice were injected intraperitoneally with 150 μ g of 145-2C11 and 24 h later with 150 μ g of H57-597 (ref. 27). Control mice were injected twice with PBS. We isolated thymocytes 48 h after the second injection and stained them for flow cytometry.

Immunisolation of TCR^{low} thymocytes

Freshly isolated thymocytes from $\delta^{+/+}$ and Δ CYT mice were resuspended in RPMI culture medium containing 2% FCS and incubated with 145-2C11 for 1 h at 4 °C, followed by two washes and incubation with mouse anti-hamster immunoglobulin (Ig) for 1 h at 4 °C. Cells were then incubated with anti-mouse Ig-coated magnetic beads (Dyna) at a 3:1 bead-to-TCR^{int/high} cell ratio. After 1 h incubation at 4 °C, we separated bead-bound cells

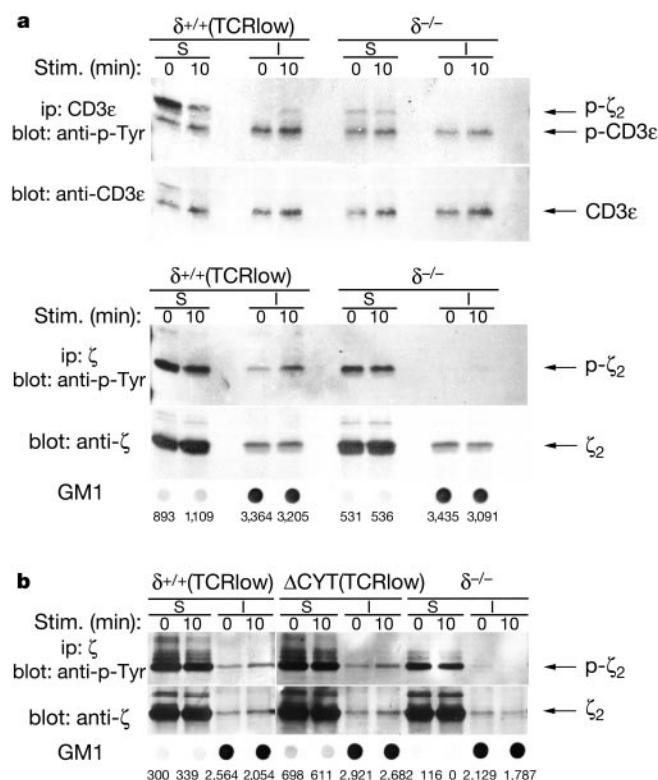


Figure 4 Impaired tyrosine phosphorylation of CD3 ζ in lipid rafts of $\delta^{-/-}$ thymocytes. Detergent-soluble (S) and -insoluble (I) fractions of anti-CD3-stimulated thymocytes were immunoprecipitated and immunoblotted with the indicated antibodies. The distribution of GM1 in each fraction is shown by dot blotting; numbers below represent densitometric

values. The positions of phospho-CD3 ϵ (p-CD3 ϵ) and phospho-CD3 ζ dimer (p- ζ_2) are indicated for $\delta^{+/+}$ (TCR^{low}) and $\delta^{-/-}$ thymocytes (**a**), and, comparatively, for Δ CYT(TCR^{low}) thymocytes (**b**).

using a magnet (Dyna). We evaluated the effectiveness of depletion of TCR^{int/high} thymocytes by flow cytometry.

TCR stimulation, immunoprecipitation and immunoblotting

Thymocytes were resuspended at 10^6 cells ml⁻¹ in RPMI. TCR stimulation was performed, unless otherwise indicated, by incubation with $10 \mu\text{g ml}^{-1}$ 145-2C11 for 10 min at 4 °C followed by crosslinking with $20 \mu\text{g ml}^{-1}$ of rabbit anti-hamster antibody at 37 °C for the times indicated. After stimulation, cells were lysed in a buffer containing 1% Brij-96, 20 mM Tris-HCl pH 7.8, 150 mM NaCl and a cocktail of protease and phosphatase inhibitors. We carried out immunoprecipitation and immunoblotting as described²⁶. For assessment of Ras-GTP, detergent lysates were incubated with a GST fusion protein containing the Ras-binding domain of Raf-1 (Raf-1 RBD agarose, Upstate Biotechnology). We detected bound Ras by immunoblotting with an anti-Ras antibody (Upstate Biotechnology).

Fractionation of plasma membrane proteins

Detergent-soluble and -insoluble (lipid rafts, glycolipid-enriched microdomains) cell membrane fractions were isolated using a modification of a discontinuous density gradient purification protocol²¹. We lysed $\delta^{-/-}$, $\delta^{+/+}$ (TCR^{low}) or ΔCYT (TCR^{low}) thymocytes (100×10^6 per stimulation point) in TNE buffer (25 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM EDTA) containing 1% Brij96 and protease and phosphatase inhibitors. Cell lysates were centrifuged at equilibrium at 2 °C on a discontinuous sucrose gradient for 15 h in a final volume of 4.5 ml, as described²¹. Fractions of 0.7 ml were collected from the top. The two first fractions, containing the detergent-insoluble proteins, were pooled and adjusted to 1% Brij96 and 60 mM octylglucoside (Roche Diagnostics). The two last fractions were also pooled and contained the detergent-soluble proteins. We assessed the effectiveness of membrane fractionation by dot blotting of ganglioside GM1 with peroxidase-conjugated cholera toxin B subunit (Sigma).

Received 2 March; accepted 22 May 2000.

- Malissen, M. *et al.* Altered T cell development in mice with a targeted mutation of the CD3- ϵ gene. *EMBO J.* **14**, 4641–4653 (1995).
- Haks, M. C., Krimpenfort, P., Borst, J. & Kruisbeek, A. M. The CD3 γ chain is essential for development of both the TCR $\alpha\beta$ and TCR $\gamma\delta$ lineages. *EMBO J.* **17**, 1871–1882 (1998).
- Dave, V. P. *et al.* CD3 δ deficiency arrest development of the $\alpha\beta$ but not the $\gamma\delta$ T cell lineage. *EMBO J.* **16**, 1360–1370 (1997).
- Swan, K. A. *et al.* Involvement of p21ras distinguishes positive and negative selection in thymocytes. *EMBO J.* **14**, 276–285 (1995).
- O'Shea, C. C., Crompton, T., Rosewell, I. A., Hayday, A. C. & Owen, M. J. Raf regulates positive selection. *Eur. J. Immunol.* **26**, 2350–2355 (1996).
- Alberola-Ila, J., Forbush, K. A., Seger, R., Krebs, E. G. & Perlmutter, R. M. Selective requirement for MAP kinase activation in thymocyte differentiation. *Nature* **373**, 620–623 (1995).
- Dave, V. P., Zdancewicz, M. & Kappes, D. J. CD4 and CD8 T cell lineages show differential sensitivity to alterations in surface density and composition of the TCR complex. *Immunity* (submitted).
- Jameson, S. C., Hogquist, K. A. & Bevan, M. J. Positive selection of thymocytes. *Annu. Rev. Immunol.* **13**, 93–126 (1995).
- D'Ambrosio, D., Cantrell, D. A., Frati, L., Santoni, A. & Testi, R. Involvement of p21ras activation in T cell CD69 expression. *Eur. J. Immunol.* **24**, 616–620 (1994).
- Page, G. *et al.* Defective thymocyte maturation in p44 MAP kinase (ERK1) knockout mice. *Science* **286**, 1374–1377 (1999).
- Sharp, L. L., Schwarz, D. A., Bott, C. M., Marshall, C. J. & Hedrick, S. M. The influence of the MAPK pathway on T cell lineage commitment. *Immunity* **7**, 609–618 (1997).
- Bomhardt, U., Basson, M. A., Krummrei, U. & Zamoyska, R. Activation of the extracellular signal-related kinase/mitogen-activated protein kinase pathway discriminates CD4 versus CD8 lineage commitment in the thymus. *J. Immunol.* **163**, 715–722 (1999).
- Schaeffer, H. J. & Weber, M. J. Mitogen-activated protein kinases: specific messages from ubiquitous messengers. *Mol. Cell. Biol.* **19**, 2435–2444 (1999).
- Rudd, C. E. Adaptors and molecular scaffolds in immune cell signalling. *Cell* **96**, 5–8 (1999).
- Clements, J. L. *et al.* Requirement for the leukocyte-specific adapter protein SLP-76 for normal T cell development. *Science* **281**, 416–419 (1998).
- Zhang, W. *et al.* Essential role of LAT in T cell development. *Immunity* **10**, 323–332 (1999).
- Bustelo, X. R. Regulatory and signaling properties of the Vav family. *Mol. Cell. Biol.* **20**, 1461–1477 (2000).
- Finco, T. S., Kadlec, T., Zhang, W., Samelson, L. E. & Weiss, A. LAT is required for TCR-mediated activation of PLC γ 1 and the Ras pathway. *Immunity* **7**, 617–626 (1998).
- Xavier, R., Brennan, T., Li, Q., McCormack, C. & Seed, B. Membrane compartmentation is required for efficient T cell activation. *Immunity* **8**, 723–732 (1998).
- Montixi, C. *et al.* Engagement of T cell receptor triggers its recruitment to low-density detergent-insoluble membrane domains. *EMBO J.* **17**, 5334–5348 (1998).
- Zhang, W., Trible, R. P. & Samelson, L. E. LAT palmitoylation: its essential role in membrane microdomain targeting and tyrosine phosphorylation during T cell activation. *Immunity* **9**, 239–246 (1998).
- Kersh, E. N., Shaw, A. S., Allen, P. M. Fidelity of T cell activation through multistep T cell receptor ζ phosphorylation. *Science* **281**, 572–575 (1998).
- Neel, B. G. Role of phosphatases in lymphocyte activation. *Curr. Opin. Immunol.* **9**, 405–420 (1997).
- Bäckström, B. T., Müller, U. & Palmer, E. Positive selection through a motif in the $\alpha\beta$ T cell receptor. *Science* **281**, 835–838 (1998).
- Malissen, B., Ardouin, L., Lin, S., Gillet, A., & Malissen, M. Function of the CD3 subunits of the pre-TCR and TCR complexes during T cell development. *Adv. Immunol.* **72**, 103–148 (1999).
- Sahuquillo, A. G., Roumier, A., Teixeira, E., Bragado, R., & Alarcón, B. T cell receptor (TCR) engagement in apoptosis-defective, but interleukin 2 (IL-2)-producing, T cells results in impaired ZAP70/CD3 ζ association. *J. Exp. Med.* **187**, 1179–1192 (1998).
- Dave, V. P. *et al.* Altered functional responsiveness of thymocyte subsets from CD3 δ -deficient mice to TCR-CD3 engagement. *Int. Immunol.* **10**, 1481–1490 (1998).

28. Dutz, J. P., Ong, C. J., Marth, J., & Teh, H. S. Distinct differentiative stages of CD4⁺CD8⁺ thymocyte development defined by the lack of coreceptor binding in positive selection. *J. Immunol.* **154**, 2588–2599 (1995).

Acknowledgements

We thank G. Koretzky and L. Samelson for antibodies; J. Millán for technical advice; J. Palacin, M. A. Bordallo, M. C. Moreno and M. Gómez for technical support; and M. Alonso and E. Bernstein for critically reading the manuscript. This work was supported by grants from Comisión Interministerial de Ciencia y Tecnología and Comunidad de Madrid. P.D. was supported by a Comunidad de Madrid fellowship and E.F. by a Ministerio de Educación fellowship. The Centro de Biología Molecular is in part supported by Fundación Ramón Areces.

Correspondence and requests for materials should be addressed to B.A. (e-mail: Balarcon@cbm.uam.es) or E.F. (e-mail: efernandez@cbm.uam.es).

Chfr defines a mitotic stress checkpoint that delays entry into metaphase

Daniel M. Scolnick*† & Thanos D. Halazonetis*

*The Wistar Institute, 3601 Spruce Street, Philadelphia, Pennsylvania 19104-4268, USA

†Graduate Program in Biochemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104-4268, USA

Chemicals that target microtubules induce mitotic stress by affecting several processes that occur during mitosis. These processes include separation of the centrosomes in prophase, alignment of the chromosomes on the spindle in metaphase and sister-chromatid separation in anaphase^{1,2}. Many human cancers are sensitive to mitotic stress. This sensitivity is being exploited for therapy and implies checkpoint defects^{2–8}. The known mitotic checkpoint genes, which prevent entry into anaphase when the chromosomes are not properly aligned on the mitotic spindle, are, however, rarely inactivated in human cancer^{9–13}. Here we describe the *chfr* gene, which is inactivated owing to lack of expression or by mutation in four out of eight human cancer cell lines examined. Normal primary cells and tumour cell lines that express wild-type *chfr* exhibited delayed entry into metaphase when centrosome separation was inhibited by mitotic stress. In contrast, the tumour cell lines that had lost *chfr* function entered metaphase without delay. Ectopic expression of wild-type *chfr* restored the cell cycle delay and increased the ability of the cells to survive mitotic stress. Thus, *chfr* defines a checkpoint that delays entry into metaphase in response to mitotic stress.

The forkhead-associated (FHA) domain was initially identified in transcription factors that have forkhead DNA-binding domains and in protein kinases¹⁴, but many proteins that contain FHA domains are cell-cycle checkpoints. To try to isolate novel mitotic checkpoint genes, we searched the expressed sequence tag (EST) database for complementary DNAs with FHA motifs. One of the identified cDNAs, corresponding to EST clones 650972 and 1071323, was sequenced and shown to encode a 664-amino-acid protein with FHA and ring finger domains within its amino terminus^{14–16} and, within its carboxy terminus, a cysteine-rich region that does not display significant similarity to any protein in the Genbank database, but is very highly conserved between humans and mice (Fig. 1a). As described below, this protein functions as a cell-cycle checkpoint; we therefore named it Chfr (checkpoint with FHA and ring finger). Chfr may be a member of a small family of proteins that contain FHA and ring finger domains. Other members of this family are Dma1 (defective in mitotic arrest 1), a mitotic checkpoint

protein in *Schizosaccharomyces pombe*¹⁷, and Yhr115c and Ynl116w, the predicted protein products of two uncharacterized *Saccaromyces cerevisiae* open reading frames (Fig. 1a). Dma1, Yhr115c and Ynl116w are highly related to each other, whereas Chfr bears less similarity to these three proteins and may not be their human orthologue (Fig. 1b, c).

Chfr expression was ubiquitous in normal human tissues (Fig. 2a), but three out of eight human cancer cell lines examined did not contain *chfr* messenger RNA (Fig. 2b). The same cell lines also did not express Chfr protein, as determined by immunoblotting with an affinity-purified polyclonal antibody (Fig. 2b; see Fig. 3c for characterization of antibody specificity). The molecular basis for the lack of *chfr* expression has not been determined, but does not involve deletion of both copies of the *chfr* gene (data not shown). Nevertheless, the high frequency of lack of detectable *chfr* expression prompted us to examine whether *chfr* is mutated in the five cancer cell lines that express its mRNA and protein. We isolated mRNA and amplified the *chfr* coding region by polymerase chain reaction with reverse transcriptase (RT-PCR) before sequencing it. U2OS was the only cell line that displayed a sequence variation (Fig. 2c). The variation consisted of a C→T transition in the non-coding strand, leading to substitution of Val 580 with Met in the highly conserved C-terminal cysteine-rich region of Chfr, and affected the entire pool of U2OS mRNA, as the sequencing did not reveal any wild-type sequence. This transition could be a mutation, because it occurred within a CG dinucleotide, and these are mutagenesis hot-spots¹⁸.

Because Chfr and Dma1 share structural domains, we examined whether Chfr is a mitotic checkpoint. The eight cancer cell lines described above were scored for their ability to undergo mitotic arrest 16 h after being treated with nocodazole, which elicits mitotic stress by inducing microtubule depolymerization. In the cell lines that had no detectable *chfr* expression and the U2OS cells, which expressed the variant *chfr* gene, the fraction of cells that had condensed chromosomes (mitotic index) was high, indicating arrest in metaphase. In contrast, the mitotic index of the cell lines that expressed wild-type *chfr* was low (Fig. 3a). To determine whether *chfr* accounted for the different response, we prepared DLD1 and U2OS cells stably expressing heamagglutinin (HA)-tagged wild-type Chfr, Chfr with the Val 580-to-Met substitution (M580) or just the *neo* selectable marker. The cells expressing *neo* or Chfr-M580 had a high mitotic index in response to nocodazole, like the parent cells; the cells expressing wild-type Chfr, however, had a low mitotic index (Fig. 3b). Near-normal levels of wild-type Chfr

protein were sufficient to affect the response to mitotic stress, as the ectopic Chfr protein in the stably transfected DLD1-*chfr* cells was expressed at levels similar to those of endogenous Chfr in primary human cells (Fig. 3c; top). Furthermore, the different effects of wild-type Chfr and Chfr-M580 could not be attributed to differences in protein expression (Fig. 3c; bottom), indicating that the nucleotide transition targeting *chfr* in U2OS cells is probably a mutation, because it leads to loss of function.

To strengthen the link between Chfr and the response to mitotic stress, we examined whether a dominant negative Chfr mutant would alter the behaviour of cells, such as SAOS2, that express wild-type Chfr and have a low mitotic index in response to mitotic stress. Chfr-ΔFHA, a Chfr protein with a deletion of residues 2–142 encompassing the FHA domain, was identified as a dominant negative mutant by studying its function in DLD1 cells (see

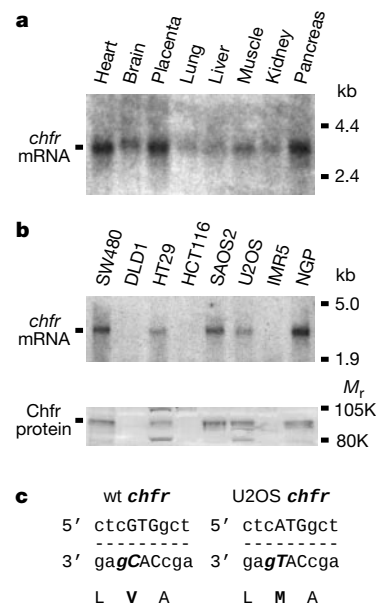


Figure 2 Expression of *chfr*. **a**, In normal human normal tissues; **b**, in cancer cell lines. M_r , relative molecular mass. **c**, Comparison of the sequences of wild-type (wt) and variant *chfr* cDNA from U2OS cells corresponding to codons 579–581. The codon for Val 580, which is mutated to Met, is shown in capital letters. The dinucleotide CG, which is mutated to TG is in bold and italicized. L, Leu; V, Val; M, Met; A, Ala.

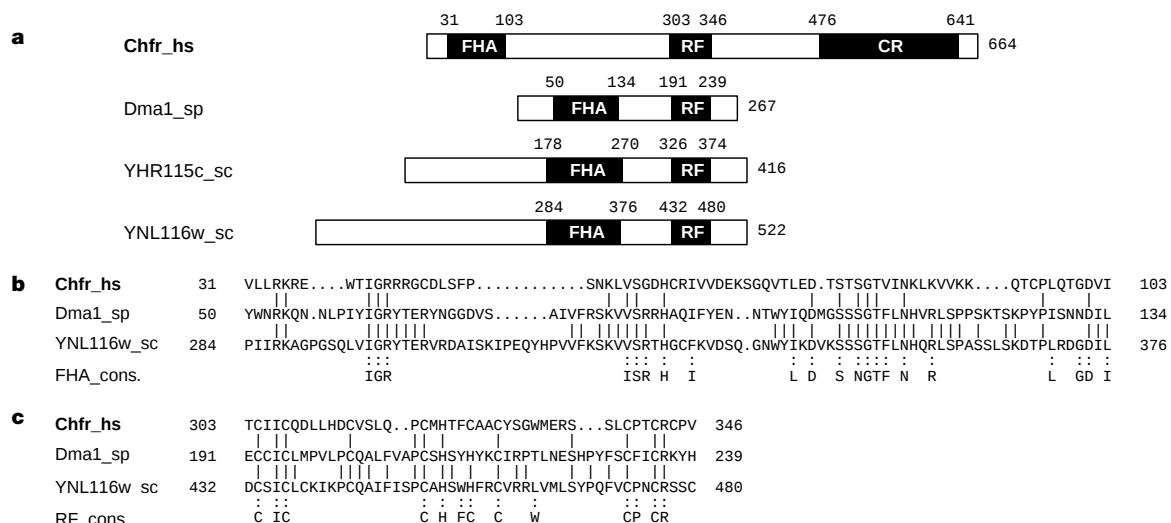


Figure 1 Proteins containing FHA and ring finger domains. **a**, Structural domains of Chfr, Dma1, Yhr115c and Ynl116w. FHA, forkhead-associated domain; RF, ring finger; CR, cysteine-rich region. The numbers refer to amino-acid positions. **b,c**, Alignment of the FHA (**b**) and ring finger (**c**) domains of human Chfr, *S. pombe* Dma1 and *S. cerevisiae*

predicted open reading frame YNL116w. The consensus (cons.) sequences of the FHA and ring finger domains are also indicated. Species: hs, human; sp, *S. pombe*; sc, *S. cerevisiae*.

below). Its effect on the response of SAOS2 cells to mitotic stress was studied by transiently transfecting these cells with plasmids that express Chfr- Δ FHA, wild-type Chfr or no Chfr protein, together with a plasmid expressing green fluorescent protein (GFP) as a marker. After 36 h we induced mitotic stress by exposure to taxol, and the mitotic index was determined 8–16 h later. About 50% of the cells expressed GFP, but the variable level of expression made it difficult to define a threshold above which a cell would be considered GFP-positive. Thus, to avoid any bias, we calculated the mitotic index for the entire cell population. Expression of wild-type Chfr had no effect as compared with empty vector (Fig. 3d). However, Chfr- Δ FHA, whose level of expression was equivalent to that of wild-type Chfr (Fig. 3e), led to a fivefold increase in the mitotic index at 12, 14 and 16 h, indicating a checkpoint defect. At the 8-h timepoint, the mitotic index was low, similar to that in cells that lack Chfr (for example, DLD1 and HCT116), which begin to show a high mitotic index in response to mitotic stress 12–16 h after addition of nocodazole or taxol (unpublished observations and ref. 9). We propose that Chfr- Δ FHA acted in this assay through dominant inhibition of endogenous wild-type Chfr, on the basis of our analysis of its function in transiently transfected DLD1 cells, which lack endogenous Chfr. On its own, Chfr- Δ FHA had no effect on the mitotic index of DLD1 cells exposed to mitotic stress, as compared to vector control, but it inhibited the ability of wild-type Chfr to decrease the mitotic index. In the same assay, Chfr-M580 did not inhibit the activity of wild-type Chfr (Fig. 3f).

The low mitotic index of nocodazole-treated cells expressing wild-type Chfr could indicate either exit from mitosis due to failure

to arrest in metaphase or arrest in the cell cycle before entry into metaphase. To distinguish between these possibilities, the stably transfected DLD1-*chfr* and DLD1-*neo* cells were synchronized by consecutive thymidine-aphidicolin blocks at the G1–S boundary and then allowed to proceed through the cell cycle with or without mitotic stress. We monitored progression through the cell cycle by measuring the mitotic index and by flow-cytometric analysis of the DNA content of the cells. In the absence of mitotic stress, Chfr had no effect on cell-cycle progression (Fig. 4a and data not shown). In the presence of mitotic stress, which was induced by adding nocodazole 12 h after release from the G1–S block, the DLD1-*neo* cells entered metaphase with the same kinetics as in the absence of mitotic stress; however, the DLD1-*chfr* cells delayed entry into metaphase by about 6 h. Similar results were obtained when mitotic stress was induced by colcemid or taxol, two other drugs that affect microtubule dynamics (Fig. 4a). The timing of induction of mitotic stress was critical for Chfr to delay entry into metaphase; when mitotic stress was induced 14 h after release from the G1–S block, both DLD1-*neo* and DLD1-*chfr* cells entered metaphase with the same kinetics as in the absence of mitotic stress (Fig. 4a). To determine that the ability of Chfr to delay metaphase entry in response to mitotic stress was not specific to DLD1 cells, we examined synchronized U2OS cells stably transfected with plasmids expressing *neo* or wild-type *chfr* (described above; see Fig. 3b). These cells behaved identically to their DLD1 counterparts (data not shown). Furthermore, human primary epidermal keratinocytes and osteoblasts also exhibited a delay in metaphase entry in response to mitotic stress (Fig. 4b).

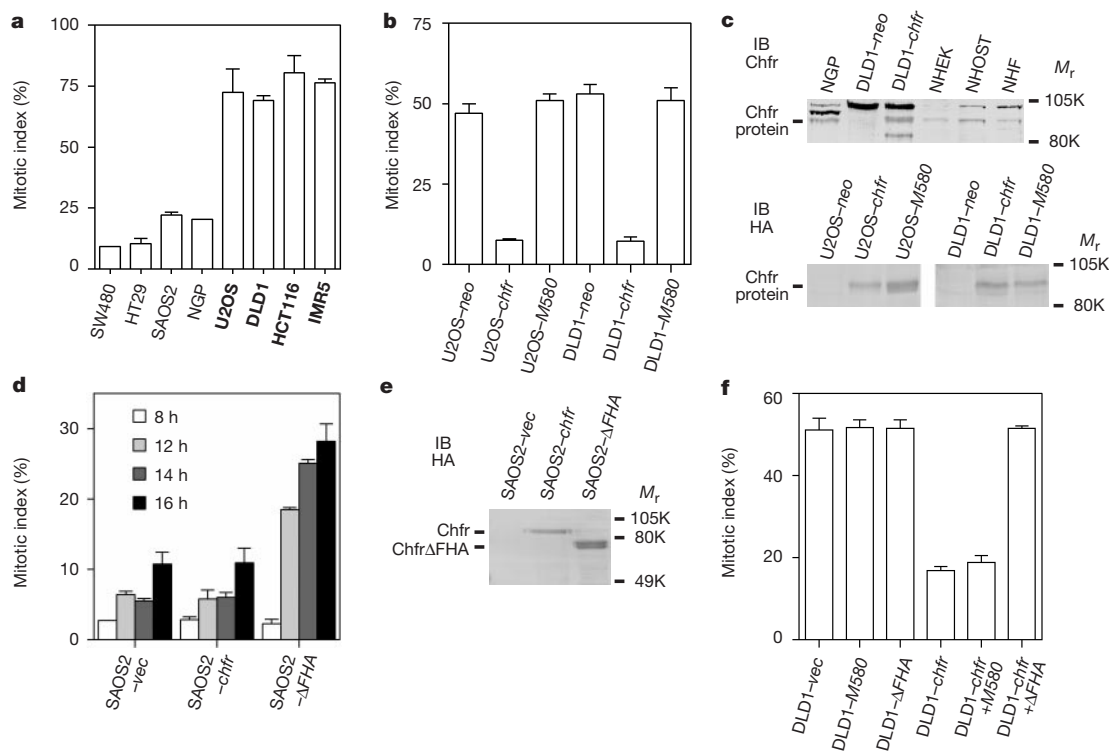


Figure 3 Chfr regulates the response of cells to mitotic stress. **a**, Mitotic index of unsynchronized human cancer cell lines after exposure to nocodazole for 16 h. The cell lines that do not express Chfr or express the variant form are in bold. **b**, Mitotic index of unsynchronized U2OS and DLD1 cells stably transfected with plasmids expressing *neo* or wild-type *chfr* or *chfr*-M580 after exposure to nocodazole for 16 h. **c**, Chfr protein levels in stably transfected DLD1 and U2OS cells expressing various HA-tagged Chfr proteins or *neo*, and levels of endogenous Chfr in NGP cells and in normal (primary) human epidermal keratinocytes (NHEK), osteoblasts (NHOST) and fibroblasts (NHF), as determined by immunoblotting (IB) with antibodies against Chfr or the HA tag. **d**, Mitotic index of unsynchronized SAOS2 cells transiently transfected with plasmids expressing no Chfr

protein (*vec*), wild-type Chfr or Chfr Δ FHA. Taxol was added 36 h after the transient transfection and the mitotic index was determined 8, 12, 14 and 16 h later. **e**, Expression of wild-type Chfr and Chfr Δ FHA in transiently transfected SAOS2 cells, as determined by immunoblotting with an antibody against their N-terminal HA tag. **f**, Mitotic index of unsynchronized DLD1 cells transiently transfected with plasmids expressing no Chfr protein (*vec*), wild-type Chfr (1 μ g), ChfrM580 (5 μ g), Chfr Δ FHA (5 μ g), wild-type Chfr (1 μ g) and ChfrM580 (5 μ g), or wild-type Chfr (1 μ g) and Chfr Δ FHA (5 μ g). Taxol was added 36 h after the transient transfection and the mitotic index was determined 16 h later.

The observation that *Chfr* delays entry of DLD1 cells into metaphase when mitotic stress was induced 12 h, but not 14 h, after release from the G1–S block implies that the function of *Chfr* is related to process(es) that occur between these timepoints. We used cyclin B/cdc2 activity and the position of the centrosomes relative to each other as markers to define the specific phases of the cell cycle corresponding to the 12-h and 14-h timepoints. Cyclin B/cdc2 activity is high in prophase and metaphase and low in G2 and anaphase^{6,7,19}, and the centrosomes are closely positioned throughout G2, but during prophase they separate and move along the periphery of the nucleus towards opposite sides of the cell^{1,20}. Twelve hours after release from the G1–S block, cyclin B/cdc2 activity was low and the centrosomes had not separated, indicating that the cells were in G2 (Fig. 5a, c). At 14 h the cells were in prophase, as cyclin B/cdc2 activity was high and centrosome separation was complete (compare with the 16-h timepoint, when cells were in metaphase). Thus, the function of *Chfr* must relate to process(es) that occur during prophase, such as separation of the centrosomes.

Centrosome separation and cyclin B/cdc2 activity were also monitored in synchronized DLD1-*neo* and DLD1-*chfr* cells treated with nocodazole 12 h after release from the G1–S block. At the 14-h timepoint, centrosome separation was inhibited in both the DLD1-*neo* and DLD1-*chfr* cells (Fig. 5b); but a significant number of DLD1-*neo* cells had condensed chromosomes, whereas the DLD1-*chfr* cells typically did not exhibit chromosome condensation (Figs 4a, 5b). Thus, *Chfr* enforced temporal dependency between centrosome separation and chromosome condensation. Cyclin B/cdc2 activity was high in synchronized DLD1-*neo* and DLD1-*chfr* cells treated with nocodazole 12 h after release from the G1–S block (Fig. 5c). Persistence of high cyclin B/cdc2 activity indicates arrest in mitosis; DLD1-*neo* cells were arrested in metaphase owing to activation of the spindle checkpoint, whereas DLD1-*chfr* cells

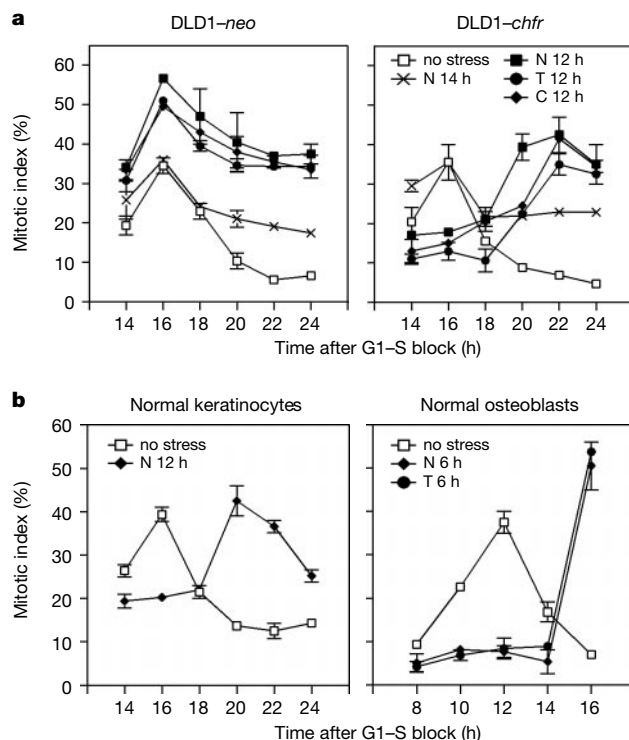


Figure 4 *Chfr* delays entry into metaphase in response to mitotic stress. **a**, Mitotic index of synchronized DLD1 cells stably transfected with plasmids expressing *neo* (DLD1-*neo*) or *chfr* (DLD1-*chfr*) as a function of time after release from the G1–S block. The cells were either not exposed to mitotic stress or treated with nocodazole (N), taxol (T) or colcemid (C) 12 or 14 h after release from the cell-cycle block. **b**, Mitotic index of synchronized normal (primary) human epidermal keratinocytes and osteoblasts in the absence and presence of mitotic stress induced 12 and 6 h, respectively, after release from the G1–S block.

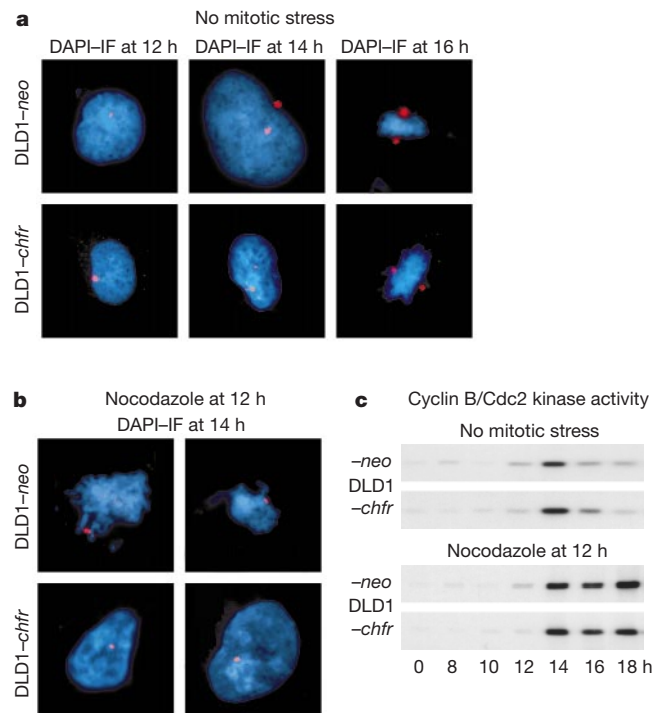


Figure 5 Centrosome separation and cyclin B/cdc2 activity in synchronized DLD1-*neo* and DLD1-*chfr* cells. **a**, Representative images of DLD1-*neo* and DLD1-*chfr* cells at 12, 14 and 16 h after release from the G1–S block in the absence of mitotic stress. DNA was stained with DAPI (blue) and centrosomes (red) were identified by immunofluorescence (IF). **b**, Disjunction of chromosome condensation and centrosome separation in cells lacking *chfr*. Representative images of DLD1-*neo* and DLD1-*chfr* cells exposed to nocodazole 12 h after release from the G1–S block and analysed by fluorescence microscopy at 14 h. **c**, Cyclin B/cdc2 kinase activity of synchronized DLD1-*neo* and DLD1-*chfr* cells as a function of time after release from the cell-cycle block without mitotic stress or with mitotic stress induced by nocodazole 12 h after release from the G1–S block.

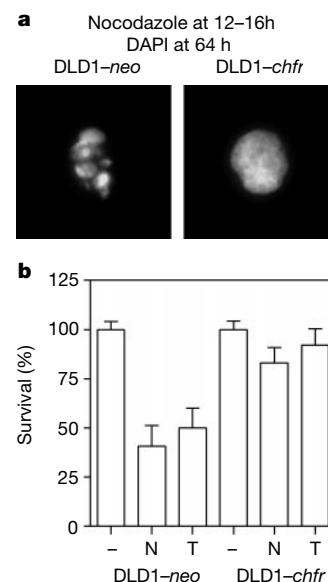


Figure 6 Response of DLD1-*neo* and DLD1-*chfr* cells to transient exposure to mitotic stress. Synchronized cells were exposed to nocodazole 12 h after release from the G1–S block for a 4-h period. **a**, Short-term response. The cells were stained with DAPI and inspected by fluorescence microscopy 64 h after release from the G1–S block. **b**, Long-term response (survival). The cells were replated and scored for colony formation three weeks later. N, nocodazole; T, taxol.

were arrested initially in prophase by the Chfr checkpoint and later in metaphase by the spindle checkpoint (Fig. 4a and data not shown). The high cyclin B/cdc2 activity in cells whose entry into metaphase is delayed by Chfr distinguishes the Chfr checkpoint from the G2 DNA-damage checkpoint, which delays entry into mitosis by inhibiting cyclin B/cdc2 (ref. 21).

To test the hypothesis that Chfr is a mitotic checkpoint further, we examined whether it affects cell viability in response to mitotic stress. Stably transfected DLD1-*neo* and DLD1-*chfr* cells were synchronized by sequential thymidine-aphidicolin blocks and exposed to nocodazole for 4 h starting 12 h after release from the G1-S block. The short-term response of the cells to mitotic stress was evaluated by examining their DNA content and nuclear morphology 48 h later. The DLD1-*chfr* cells exhibited the normal DNA content profile of cycling cells and normal nuclear morphology (Fig. 6a and data not shown). The DNA content profile of the DLD1-*neo* cells was also normal, but their nuclear morphology was clearly aberrant; about half of all cells with a 4N DNA content exhibited fragmented nuclei, indicating that they may not have completed mitosis properly (Fig. 6a). To study the effect of mitotic stress on cell viability, synchronized DLD1-*neo* and DLD1-*chfr* cells were transiently exposed to nocodazole or taxol, as described above, and then allowed to form colonies over a three-week period. DLD1-*neo* cells showed a decrease in the number of colony-forming units (CFUs) in response to mitotic stress, whereas for the DLD1-*chfr* cells the number of CFUs was unaffected by mitotic stress (Fig. 6b).

Most of the research on mitotic checkpoints has focused on the spindle checkpoint, which monitors the transition from metaphase to anaphase^{6,7}. Although this transition is clearly important, other phases of mitosis are also likely to be monitored by checkpoints. A checkpoint monitoring the anaphase-to-telophase transition has been described²². The *chfr* gene molecularly defines the existence of a checkpoint that regulates entry into metaphase. The Chfr checkpoint was evident in primary human cells, but was inactivated in four out of eight human cancer cell lines examined. In the absence of the Chfr checkpoint, cells subjected to mitotic stress condensed their chromosomes despite failing to separate their centrosomes. Many questions remain unanswered regarding the function of Chfr. Our current model, that Chfr monitors centrosome separation rather than some other mitotic stress-sensitive event that occurs in prophase, needs to be confirmed; the molecular mechanism by which Chfr delays cell-cycle progression needs to be determined; and the frequency of Chfr inactivation in primary tumours needs to be established. So far, analysis of a small number of cancer cell lines raises the possibility that Chfr is inactivated more frequently than all known spindle checkpoint genes combined⁹⁻¹³. If Chfr is inactivated in human cancer, then its inactivation may underlie the increased sensitivity of cancer cells to antimitotic drugs. □

Methods

Chfr expression in normal tissues and cancer cell lines

We analysed expression of *chfr* mRNA using a human multiple tissue northern blot (Clontech) and northern blots of mRNA isolated from cancer cell lines. For analysis of Chfr protein expression, whole cell lysates²³ were immunoblotted with an affinity-purified rabbit polyclonal antibody prepared using histidine-tagged Chfr as the antigen (Research Genetics).

Cell culture

All cancer cell lines were grown in 10% fetal bovine serum/DMEM; normal human epidermal keratinocytes and osteoblasts were grown in KGM2 and OGM media, respectively (Clonetics). For synchronization, the cells were treated with 2 mM thymidine for 16 h, then with 0.25 mM thymidine/deoxycytidine for 9 h and then with 5 µg ml⁻¹ aphidicolin for 20–24 h (ref. 24). To induce mitotic stress, cells were exposed to 0.5 µg ml⁻¹ nocodazole, 5 µM taxol or 0.5 µg ml⁻¹ colcemid.

Ectopic Chfr expression

pSV2-HA-*chfr* contains an insert encoding full-length Chfr fused at its N terminus to an HA tag in the pSV2 vector²³. pSV2-HA-*chfr*-M580 and pSV2-HA-*chfr*-ΔFHA were

derived from pSV2-HA-*chfr* by site-directed mutagenesis. We transfected DLD1 or U2OS cells with 5 µg of one of the pSV2-HA-*chfr* vectors or pSV2 vector without insert and 1 µg pSV7neo plasmid using Fugene-6 transfection reagent (Roche); stably transfected cells were selected with G418 (ref. 23). SAOS2 cells were transfected as described above, except that the plasmids expressing Chfr were co-transfected with a plasmid expressing GFP.

Mitotic index and centrosome staining

Cells were grown on eight-well culture slides coated with human fibronectin (Becton Dickinson). They were fixed with 1% paraformaldehyde in 0.5× KM buffer (10 mM MES (pH 6.2), 10 mM NaCl, 1.5 mM MgCl₂, 2.5% glycerol) for 15 min, washed once with 0.2% Triton X-100 in phosphate buffer saline (PBS) for 20 min and then three times with PBS. Centrosomes were stained with autoimmune serum Ab598 and DNA was stained with 4,6-diamidino-2-phenylindole (DAPI).

Viability in response to mitotic stress

Synchronized cells were either not exposed to mitotic stress or exposed to nocodazole or taxol for 4 h starting 12 h after release from the G1-S block. We evaluated the short-term response to mitotic stress 48 h after removal of nocodazole or taxol; the nuclear morphology of the cells was visualized by fluorescence microscopy and their DNA content profile was determined by flow cytometry. The long-term response to mitotic stress was evaluated by replating the cells after exposure to mitotic stress at a density of 200 cells per 100-mm diameter tissue culture dish and counting the number of colonies three weeks later.

Cdc2 kinase activity

Whole cell extracts²³ were incubated with anti-cyclin B antibody (Santa Cruz) coupled to protein G beads (Pharmacia) for 1 h. The beads were washed three times with cell lysis buffer and twice with cdc2 kinase (CK) buffer (50 mM HEPES (pH 7), 10 mM MgCl₂, 10 mM MnCl₂, 200 mM NaCl), and then incubated with 1 µg histone H1 (Upstate Biotechnology) in CK buffer supplemented with 30 mM DTT, 0.06 µM ATP and 1 µCi ³²P-γ-ATP for 20 min at 30 °C, at which time the reactions were subjected to denaturing gel electrophoresis and autoradiography.

Detection of *chfr* mutations in cancer cell lines

Messenger RNA isolated from cancer cell lines (Quickprep Micro mRNA Purification Kit, Pharmacia) was used as template for first-strand cDNA synthesis (Retroscript, Ambion). We used synthetic oligonucleotides to amplify regions of the *chfr* cDNA by PCR (Platinum Taq, Life Sciences or GC-rich PCR system, Roche). The amplified regions spanning the entire *chfr* coding sequence were sequenced using four-colour fluorescent dideoxy terminators (Big Dyes, Perkin Elmer).

Received 16 December 1999; accepted 10 May 2000.

- McIntosh, J. R. & Koonce, M. P. Mitosis. *Science* **246**, 622–628 (1989).
- Jordan, M. A. & Wilson, L. Microtubules and actin filaments: dynamic targets for cancer chemotherapy. *Curr. Opin. Cell Biol.* **10**, 123–130 (1998).
- Hartwell, L. H. & Kastan, M. B. Cell cycle control and cancer. *Science* **266**, 1821–1828 (1994).
- Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386**, 623–627 (1997).
- Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643–649 (1998).
- Elledge, S. J. Mitotic arrest: Mad2 prevents sleepy from waking up the APC. *Science* **279**, 999–1000 (1998).
- Amon, A. The spindle checkpoint. *Curr. Opin. Genet. Dev.* **9**, 69–75 (1999).
- Li, Y. & Benezra, R. Identification of a human mitotic checkpoint gene: hSMAD2. *Science* **274**, 246–248 (1996).
- Cahill, D. P. *et al.* Mutations of mitotic checkpoint genes in human cancers. *Nature* **392**, 300–303 (1998).
- Cahill, D. P. *et al.* Characterization of MAD2B and other mitotic spindle checkpoint genes. *Genomics* **58**, 181–187 (1999).
- Yamaguchi, K., Okami, K., Hibi, K., Wehage, S. L., Jen, J. & Sidransky, D. Mutation analysis of hBUB1 in aneuploid HNSCC and lung cancer cell lines. *Cancer Lett.* **139**, 183–187 (1999).
- Jin, D. Y., Spencer, F. & Jeang, K. T. Human T cell leukemia virus type 1 oncoprotein Tax targets the human mitotic checkpoint protein MAD1. *Cell* **93**, 81–91 (1998).
- Zou, H., McGarry, T. J., Bernal, T. & Kirschner, M. W. Identification of a vertebrate sister-chromatid separation inhibitor involved in transformation and tumorigenesis. *Science* **285**, 418–422 (1999).
- Hofmann, K. & Bucher, P. The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. *Trends Biochem. Sci.* **20**, 347–349 (1995).
- Lovering, R. *et al.* Identification and preliminary characterization of a protein motif related to the zinc finger. *Proc. Natl. Acad. Sci. USA* **90**, 2112–2116 (1993).
- Borden, K. L. *et al.* The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. *EMBO J.* **14**, 1532–1541 (1995).
- Murone, M. & Simanis, V. The fission yeast *dma1* gene is a component of the spindle assembly checkpoint, required to prevent septum formation and premature exit from mitosis if spindle function is compromised. *EMBO J.* **15**, 6605–6616 (1996).
- Holliday, R. & Grigg, G. W. DNA methylation and mutation. *Mutat. Res.* **285**, 61–67 (1993).
- Furuno, N., den Elzen, N. & Pines, J. Human cyclin A is required for mitosis until mid prophase. *J. Cell Biol.* **147**, 295–306 (1999).
- Zimmerman, W., Sparks, C. A. & Dosey, S. J. Amorphous no longer: the centrosome comes into focus. *Curr. Opin. Cell Biol.* **11**, 122–128 (1999).
- Weinert, T. DNA damage and checkpoint pathways: molecular anatomy and interactions with repair. *Cell* **94**, 555–558 (1998).
- Muhua, L., Adams, N. R., Murphy, M. D., Shields, C. R. & Cooper, J. A. A cytokinesis checkpoint requiring the yeast homologue of an APC-binding protein. *Nature* **393**, 487–491 (1998).

23. Chehab, N. H., Malikzay, A., Stavridi, E. S. & Halazonetis, T. D. Phosphorylation of Ser-20 mediates stabilization of human p53 in response to DNA damage. *Proc. Natl Acad. Sci. USA* **96**, 13777–13782 (1999).
24. Janss, A. J. *et al.* Caffeine and staurosporine enhance the cytotoxicity of cisplatin and camptothecin in human brain tumor cell lines. *Exp. Cell Res.* **243**, 29–38 (1998).

Acknowledgements

We thank G. Maul for the gift of autoimmune serum Ab598; A. Janss for advice on cell synchronization; J. Bothos, M. Summers, M. Appel and M. Shayhorn for technical help; M. Scolnick for editing the manuscript; and E. Stavridi, N. Chehab, G. Rovera, F. Rauscher III and P. Leder for comments and advice. We also thank the Wistar Institute's NCI-supported Microscopy, Sequencing and Bioinformatics Facilities for technical help and access to their equipment. D.M.S. is supported by an NCI Training Grant awarded to the Wistar Institute.

Correspondence and requests for materials should be addressed to T.D.H. (e-mail: halazonetis@wistar.upenn.edu). The accession number for human *chfr* is AF170724.

Humanized xenobiotic response in mice expressing nuclear receptor SXR

Wen Xie*, Joyce L. Barwick†, Michael Downes*, Bruce Blumberg‡, Cynthia M. Simon*, Michael C. Nelson*, Brent A. Neuschwander-Tetri§, Elizabeth M. Brunt||, Philip S. Guzelian† & Ronald M. Evans*

* Howard Hughes Medical Institute, Gene Expression Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

† Medical Toxicology, University of Colorado Health Sciences Centre, Denver, Colorado 80262, USA

‡ Department of Developmental and Cell Biology, University of California, Irvine, California 92697, USA

§ Division of Gastroenterology and Hepatology, and || Department of Pathology, St. Louis University School of Medicine, St. Louis, Missouri 64108, USA

The cytochrome CYP3A gene products, expressed in mammalian liver, are essential for the metabolism of lipophilic substrates, including endogenous steroid hormones and prescription drugs^{1,2}. CYP3A enzymes are extremely versatile and are inducible by many of their natural and xenobiotic substrates. Consequently, they form the molecular basis for many clinical drug–drug interactions³. The induction of CYP3A enzymes is species-specific^{4,5}, and we have postulated that it involves one or more cellular factors, or receptor-like xeno-sensors⁶. Here we identify one such factor unequivocally as the nuclear receptor pregnenolone X receptor (PXR)^{7,8} and its human homologue, steroid and xenobiotic receptor (SXR)^{8–10}. We show that targeted disruption of the mouse PXR gene abolishes induction of CYP3A by prototypic inducers such as dexamethasone or pregnenolone-16 α -carbonitrile. In transgenic mice, an activated form of SXR causes constitutive upregulation of CYP3A gene expression and enhanced protection against toxic xenobiotic compounds. Furthermore, we show that the species origin of the receptor, rather than the promoter structure of CYP3A genes, dictates the species-specific pattern of CYP3A inducibility. Thus, we can generate ‘humanized’ transgenic mice that are responsive to human-specific inducers such as the antibiotic rifampicin. We conclude that SXR/PXR genes encode the primary species-specific xeno-sensors that mediate the adaptive hepatic response, and may represent the critical biochemical mechanism of human xenoprotection.

To examine the significance of PXR in xenoregulation *in vivo*, we generated PXR-null mice with the strategy outlined in Fig. 1a. The resulting mutant allele has a deletion of two exons including amino-acid residues 63–170 of the DNA-binding domain⁷. The disruption

of PXR alleles was confirmed by the absence of PXR expression in the liver and small intestine, two principal PXR-expressing tissues (Fig. 1c). The PXR-null mice are both viable and fertile. Loss of PXR does not alter the basal expression of CYP3A (Fig. 1c). However, the CYP3A gene is no longer induced in response to the prototypic rodent-specific CYP3A inducers, pregnenolone-16 α -carbonitrile (PCN; Fig. 1d; ref. 11) and dexamethasone (DEX; Fig. 1e; refs 11, 12). This establishes that PXR is an essential mediator of CYP3A xenoregulation *in vivo*. This loss of inducibility is restricted to PXR target genes such as CYP3A. For example, the hepatic *CYP1A2* gene in the PXR-null mice remains responsive to 3-methylcholanthrene (3MC; Fig. 1e), and DEX still induces hepatic tyrosine aminotransferase, a glucocorticoid receptor target gene (data not shown).

Evolutionary divergence has been proposed to be responsible for the marked differences in drug response between humans and rodents^{6–8,13–16}. To test the ability of SXR and PXR to confer species-specific inducibility of CYP3A, we transfected the SXR gene into cultures of primary rat hepatocytes and examined the effects of a panel of steroid and nonsteroid inducers on expression from the CYP3A23 (rat) or CYP3A4 (human) promoters. We used primary cultures because the natural CYP3A promoters we used appear to be completely inactive or unresponsive in essentially all cultured cell lines.

In control rat cells without SXR, CYP3A23 was strongly induced by PCN, nifedipine or RU486, but rifampicin (RIF), clotrimazole (CTZ), phenobarbital, 3MC, corticosterone, coumestrol, cortisol, 17 β -estradiol (E2), progesterone, pregnenolone and cortisone produced little or no induction (Fig. 2a; ref. 11). No PXR vector was added, so this profile reflects the activity of endogenous PXR. In contrast, co-transfection of SXR results in significant induction of CYP3A23 by drugs known to be active in humans including RIF, CTZ, phenobarbital, E2 and pregnenolone. In addition, the induction of CYP3A23 by nifedipine and RU486 increased significantly in the presence of SXR, but activation by PCN remained unchanged (Fig. 2a). Therefore, transfection of SXR is sufficient to convert the induction response characteristics of the hepatocytes from rat to human. The critical factor is SXR, not the target CYP3A genes, as we observed the same conversion when the rat cells were co-transfected with the human CYP3A4 promoter (Fig. 2b). Indeed, we noted that when endogenous PXR is activated it induces both the rodent (Fig. 2a, lane 2) and human CYP3A genes (Fig. 2b, lane 2). SXR also activates both the human (Fig. 2b, lane 3) and rodent (Fig. 2a, lane 3) CYP3A genes in a ligand-dependent manner.

We used these assays to examine the roles of the putative SXR and PXR response elements in the activation of CYP3A promoters. The rat CYP3A23 response element is formed by two consensus half-sites organized as a direct repeat separated by 3 nucleotides (DR3)^{6–10}. Mutating either one (DR3/M2) or both (DR3/M1) half-sites abolished the PXR- and SXR-mediated activation of CYP3A23 by PCN, RIF and CTZ. In contrast, replacement of the DR3 element by an inverted repeat (IR6) element from the human CYP3A4 promoter^{6–10} rescued inducibility by PCN, RIF and CTZ (Fig. 2c). The results demonstrate that SXR/PXR and their response elements are essential in determining patterns of CYP3A inducibility.

To generate ‘humanized’ animal models of CYP3A xenoregulation, and to establish the role SXR/PXR activation in xenoprotection, we next prepared transgenic mice with expression of either wild-type SXR or an activated form of SXR (VPSXR) directed by the liver-specific albumin promoter¹⁷ (Fig. 3a). VPSXR shares similar DNA-binding specificity to SXR, and was generated by fusing the VP16 activation domain of the herpes simplex virus to the amino terminus of SXR. Expression of these constructs was confirmed by northern blotting (Fig. 3b and d). The Alb–SXR transgene was subsequently bred into a PXR-null background, and the resulting PXR-null/SXR-transgenic mice lacked mouse PXR gene but had human SXR transgene (Fig. 3e).

In the transgenic mice, a single dose of RIF (5 mg kg⁻¹) caused robust induction of mouse CYP3A11 messenger RNA in the liver of Alb-SXR mice (Fig. 3b), although no such induction occurs in wild-type mice¹⁸. In agreement with transfection results (Fig. 2), CTZ produced a higher amount of CYP3A11 in Alb-SXR mice, and PCN induced CYP3A11 equally in wild-type and transgenic mice (Fig. 3b). The induction of CYP3A in Alb-SXR mice was ligand dependent (Fig. 3b, compare lanes 6 and 7). Induction of CYP3A by RIF was rapid, near maximal by 12 h (Fig. 3c, lane 3), and was maintained for at least 3 days in the continuous presence of inducer (Fig. 3c, lanes 4 and 5). Only basal CYP3A expression was observed in nontransgenic mice, even after 7 days of treatment (Fig. 3c, lane 6). Moreover, RIF induction was reversible, as CYP3A expression fell 5 days after withdrawal from an initial 7-day treatment (Fig. 3c, lane 7). The CYP3A induction was also dose dependent (Fig. 3c, lanes 8–11). The dynamics and the reversibility of rodent CYP3A11 mRNA changes in transgenic animals are remarkably consistent with changes in enzyme levels in humans¹⁹. Thus, the expression of human SXR allows animals to respond to human-specific inducers such as RIF, consistent with our results in transfected hepatocytes.

In Alb-VPSXR mice, the *CYP3A* gene was constitutively induced in a liver- and CYP3A-specific manner (Fig. 3d, lane 5). The expression of CYP3A11 in the small intestine remained unchanged (Fig. 3d, lane 4), and neither CYP7A (Fig. 3d, compare lanes 1 and 5), nor CYP1A2 (data not shown) was induced. In addition, the Alb-VPSXR mice exhibited growth retardation, hepatomegaly and histologic liver toxicity, although the untreated SXR mice appear normal (data not shown).

To investigate whether PXR is necessary and SXR is sufficient for CYP3A induction, we examined the regulation of *CYP3A* in PXR-null/SXR-transgenic mice. As predicted, the PXR-null mice do not respond to PCN even in presence of SXR (Fig. 3e, lane 5). In contrast, the *CYP3A* gene was readily induced by the human-specific inducer RIF in the same PXR-null/SXR-transgenic mice (Fig. 3e, lane 4). Therefore, unlike their SXR direct transgenic counterparts which respond to both human- and rodent-specific inducers (Fig. 3b), the PXR-null/SXR-transgenic mice respond only to human-specific inducers such as RIF. We conclude that SXR alone can reconstitute *CYP3A* xenoregulation and confer a human profile of inducibility to PXR-deficient mice.

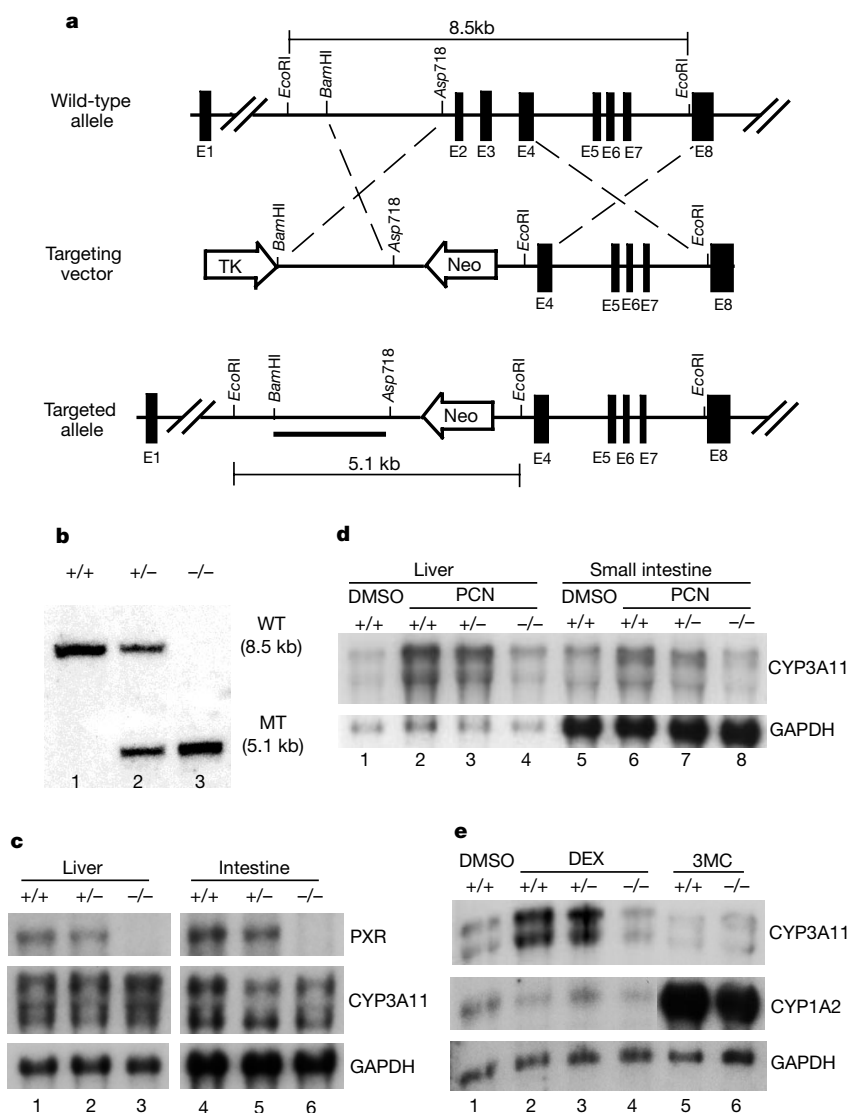


Figure 1 Targeting of the PXR gene and specific loss of xenoregulation of CYP3A in PXR null mice. **a**, Restriction map of the PXR gene and strategy to generate PXR mutant allele. The PXR probes used for Southern blotting and expected fragment sizes after *EcoRI* digestion are indicated. E, exon. Neo, neomycin resistance gene; TK, thymidine kinase promoter. **b**, Southern blotting analysis of *EcoRI*-digested genomic DNA. WT, wild type; MT, mutant. **c**, Absence of PXR mRNA in PXR^{-/-} mice. Total RNA was analysed for PXR transcripts using a PXR cDNA probe. The membrane was subsequently stripped and rehybridized with probes for CYP3A11 and GAPDH. **d**, Loss of CYP3A induction by PCN in PXR null mice. **e**, Loss of CYP3A induction by DEX, but not CYP1A2 induction by 3MC in PXR null mice.

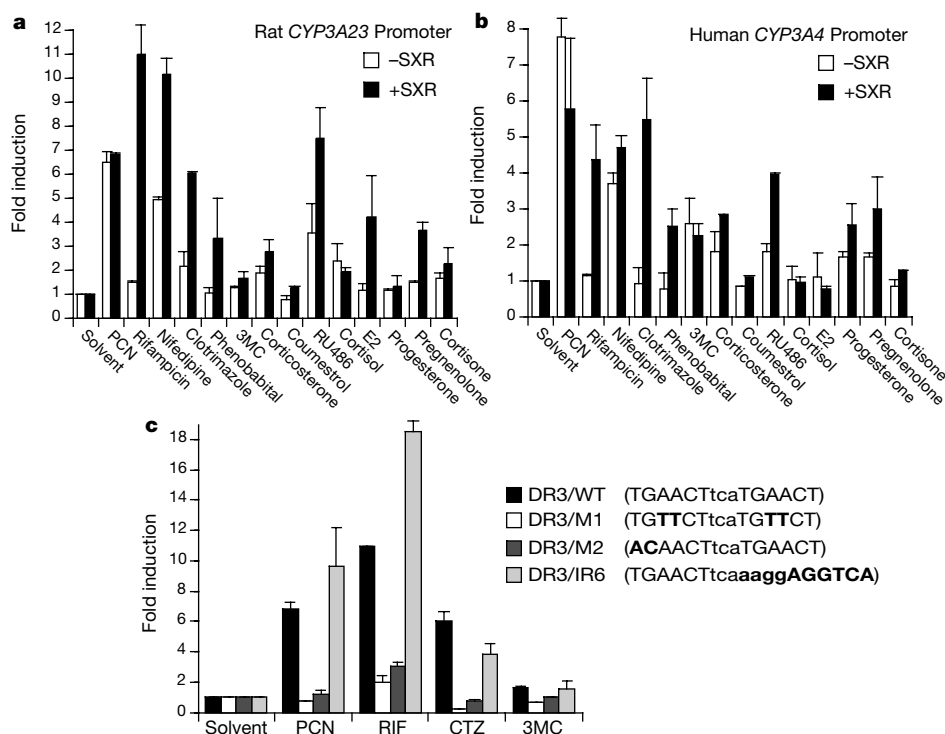


Figure 2 SXR confers human response to *CYP3A* genes in cultured primary hepatocytes. **a**, The rat *CYP3A23* cellular promoter reporter was transfected in the absence or presence of expression vector for SXR. Cells were subsequently treated with the indicated compounds. Results are shown as fold induction over solvent (DMSO), and represent the averages and standard error from triplicate assays. **b**, Similar transfections as in **a** except

that the human *CYP3A4* cellular promoter reporter was used. **c**, The DR3 element is essential for SXR-mediated activation of *CYP3A23*, and is interchangeable with the human IR6 element. The wild-type (DR3/WT) or mutant forms (DR3/M1, DR3/M2 and DR3/IR6) of *CYP3A23* cellular promoter reporters were co-transfected with SXR.

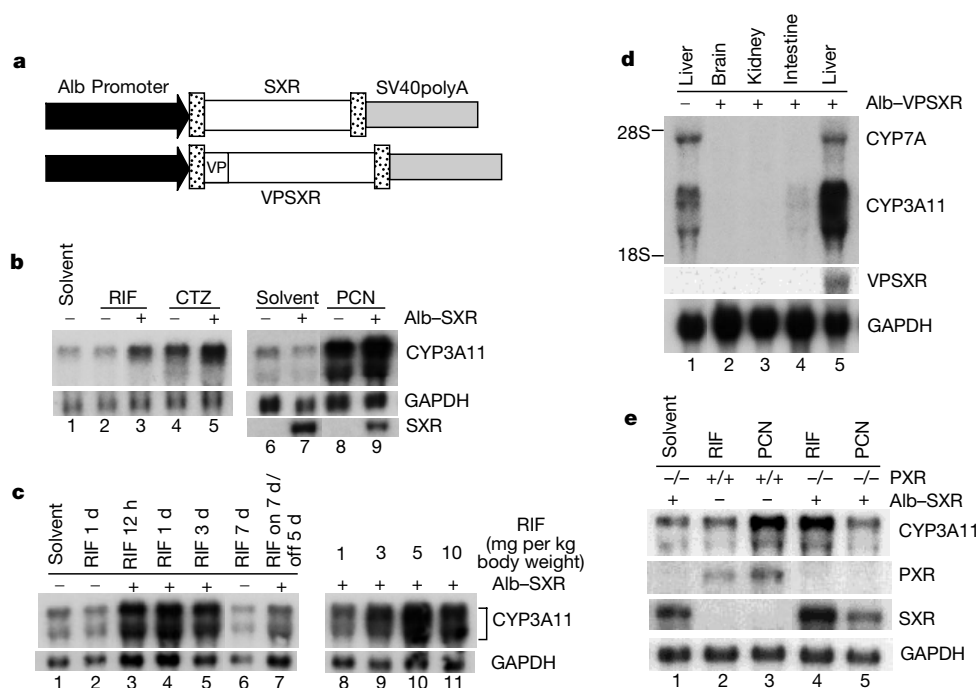


Figure 3 Generation and CYP3A regulation of SXR transgenic and PXR-null/SXR-transgenic mice. **a**, Schematic representations of the Alb-SXR and Alb-VPSXR transgene constructs. **b–e**, Northern blot analysis of liver total RNAs with the exception of the indicated tissues in **d**. **b**, Mice were treated with a single dose of RIF (5 mg kg⁻¹), CTZ or PCN. Membranes were probed for CYP3A11. **c**, Dynamics (lanes 1–7) and dose response (lanes 8–11) of RIF treatment in SXR transgenics. Mice in lanes 1–7 were

subjected to daily treatment of RIF (5 mg kg⁻¹) for the indicated period of time. In lanes 8–11, mice were treated with a single indicated dose of RIF. **d**, Constitutive and liver-specific upregulation of *CYP3A* gene in VPSXR transgenics. The membrane was probed for CYP3A11 and CYP7A. **e**, CYP3A regulation in PXR null/SXR transgenic mice. The drug treatment was as in **b**.

The sustained hepatic induction of CYP3A in Alb-VPSXR mice allowed us to explore their potential resistance to xenobiotic toxicants such as tribromoethanol and zoxazolamine. Resistance to such toxicants is a hallmark of an activated xenobiotic system and a traditional test for drug-drug interaction^{2,20}. Indeed, tribromoethanol and zoxazolamine efficiently induced anesthesia and paralysis in wild-type mice as expected²⁰. The wild-type mice slept for an average of 32.8 min in response to tribromoethanol, but the VPSXR mice were virtually resistant to the drug (Fig. 4a). Similarly with zoxazolamine treatment, wild-type mice averaged more than 60 min of paralysis, and the transgenics averaged less than 5 min (Fig. 4b). Therefore, sustained activation of SXR and induction of CYP3A result in enhanced protection against xenotoxic compounds.

Our results in hepatocyte cultures and in transgenic mice establish PXR/SXR as the central mediator for inducible expression of CYP3A. Given the widespread implications of CYP3A induction on drug and steroid metabolism, and the species-specific variation, the conversion of response profile in SXR-transgenic mice provides direct evidence that nuclear receptors can serve as xeno-sensors, and firmly establishes the molecular basis of pharmacological responses of the CYP3A gene. The factors responsible for variation in CYP3A expression between people are unclear. This variation is believed to influence the therapeutic index for up to one-third of all drugs, and may also contribute to inter-individual differences in the generation and elimination of environmental toxins and carcinogens²¹.

The extent to which drugs such as RIF can upregulate CYP3A is of pharmacological importance because, in principle this activation will not only affect RIF metabolism, but also the metabolism of any compounds processed by cytochromes^{22,23}. There has been no reliable system other than in humans to assess this problem directly and quantitatively. Thus, the generation of Alb-SXR transgenic mice is a big step toward generating a humanized rodent toxicological model. These mice readily respond to human inducers such as RIF in the equivalent range of the standard oral dosing regimen in humans (300–600 mg per 70-kg man), and exhibit similar dynamics¹⁹. Finally, the PXR-null/SXR-transgenic mice respond faithfully to human-specific inducers. This is one of the rare examples where replacing a single transcriptional regulator allows a conversion of species-specific gene regulation. Moreover, the exclusive human profile of CYP3A inducibility exhibited in the PXR-null/SXR-mice represents additional advantages for their potential usage in pharmacological studies and pharmaceutical development. Considering the wide-spread problem of drug-drug interactions and the inherent unpredictability of this process, coupled with the potential for liver toxicity as implicated in the

VPSXR mice (data not shown), there is little obvious benefit in any drug that additionally or spuriously activates SXR. Thus, screening compounds for SXR neutrality may be judicious during the drug development process. The Alb-SXR and PXR-null/SXR-transgenic mice, and the hepatocyte transfection system will be invaluable tools in such applications. □

Methods

Generation of PXR-null mice

Mouse PXR genomic DNA was isolated by screening a 129/Sv library (Stratagene) using a PXR complementary DNA probe⁸. A targeting vector was generated by replacing the second and third exons of PXR with a PGK-Neo selection marker, in conjunction with a negative selection marker (PGK-TK). After transfection, single J1 ES cell clones resistant to G418 (200 mg ml⁻¹) and ganciclovir (0.2 µM) were screened for designated homologous recombination by Southern blotting. PXR^{-/-} embryonic stem (ES) cells were micro-injected into C57BL6/J blastocysts, which were transplanted into the uteri of pseudo-pregnant ICR mice. Chimaeric male progeny were crossed with C57BL6/J females. Germline transmission of the disrupted allele was detected in agouti progeny by Southern blot and polymerase chain reaction (PCR) analysis. Mice were handled in an accredited Institute facility in accordance with the institutional animal care policies.

Plasmid constructs and mutagenesis

The CYP3A23 cellular promoter reporter, PGL3-CYP3A23, was cloned by inserting the PCR-amplified 5' regulatory sequence of rat CYP3A23 gene (nucleotides -1,360 to +82; ref. 24) into the PGL3 vector (Promega). PGL3-CYP3A4 contains up to nucleotide -1093 of the 5' flanking regions of the human CYP3A4 gene²⁵. Site-directed mutagenesis of the CYP3A23 promoter was performed by the PCR overextension method²⁶, and confirmed by DNA sequencing. The expression vectors for SXR, VPSXR and PXR were described previously⁸.

Preparation of hepatocytes, DNA transfections and drug treatment

Primary cultures of rat hepatocytes and Lipofectin (Gibco BRL)-mediated DNA transfections were carried out as described⁶. When necessary, cell were treated with RIF, DEX, PCN, nifedipine, CTZ, corticosterone, coumestrol, RU486, cortisol, E2, pregnenolone, progesterone, cortisone (10 µM (micromolar) each), phenobarbital, 3MC (2 mM each), or the control solvent. PCN was a gift from J. Babcock, other compounds were purchased from Sigma.

Generation of transgenic mice

The SXR or VPSXR cDNA were cloned downstream of the mouse albumin promoter/enhancer¹⁷. A SV40 intron/poly (A) sequence²⁷ was subsequently placed downstream of SXR and VPSXR cDNAs. Transgenes were excised and purified from the vector before microinjection into single-cell CB6F1 mouse zygotes. Transgene positive mice were screened by PCR. The Alb-SXR transgene was subsequently bred into PXR-null background to generate PXR-null/SXR-transgenic mice.

Animal drug treatment

Animals were maintained *ad libitum*. RIF was given by gastric gavage. When necessary, mice were subjected to a single intraperitoneal injection of DEX (50 mg kg⁻¹), PCN (40 mg kg⁻¹), CTZ (50 mg kg⁻¹), or 3MC (4 mg kg⁻¹) 24 h before sacrifice.

Northern blot analysis

Total RNA was prepared from tissues using TRIZOL Reagent (Gibco BRL). Northern hybridization was carried out as described²⁷. The probes of CYP3A11 (nucleotides 1,065–1,569; ref. 28), CYP7A (nucleotides 973–1,453; ref. 29), CYP1A2 (nucleotides 151–1,565; ref. 30) were cloned by PCR followed by reverse transcription from wild-type mouse liver mRNA.

Tribromoethanol anesthesia and zoxazolamine paralysis test

Modified from methods described²⁰. Six- to seven-month-old mice were used for tribromoethanol anesthesia test at 375 mg kg⁻¹ by subcutaneous injection. The sleeping mice were placed on their backs, and sleep time was measured until the animals had regained enough consciousness to right itself repeatedly. Five- to six-week-old animal were used for zoxazolamine paralysis test at 112.5 mg kg⁻¹ by intraperitoneal injection. The statistic analysis was performed with InStat 2.03 software, we used the nonparametric Mann-Whitney Test.

Received 27 March; accepted 19 June 2000.

- Gonzalez, F. J. Human cytochromes P450: problems and prospects. *Trends Pharmacol. Sci.* **13**, 346–352 (1992).
- Denison, M. S. & Whitlock, J. P. Xenobiotic-inducible transcription of cytochrome P450 genes. *J. Biol. Chem.* **270**, 18175–18178 (1995).
- Maurel, P. in *Cytochrome P450: Metabolic and Toxicological Aspects*. (ed. Ioannides, C.) 241–270 (CRC, Boca Raton, Florida, 1996).
- Gonzalez, F. J. Molecular genetics of the P-450 superfamily. *Pharmacol. Ther.* **45**, 11–38 (1990).
- Nelson, D. R. Cytochrome P450 and the individuality of species. *Arch. Biochem. Biophys.* **369**, 1–10 (1999).

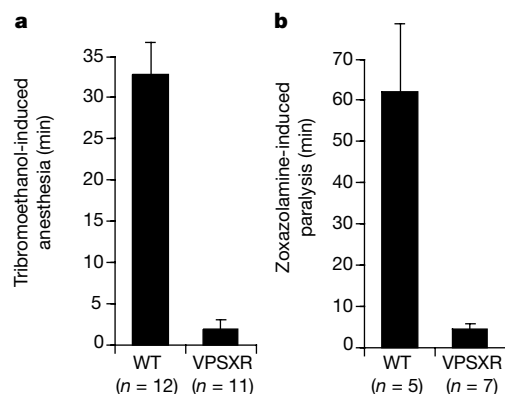


Figure 4 Enhanced protection against xenobiotic toxicants in VPSXR mice.

a, Tribromoethanol anesthesia test in males. Results represent the averages and standard error from the indicated number of mice. The two-tailed *P* value of nonparametric Mann-Whitney Test is less than 0.0001. **b**, Zoxazolamine paralysis test in males. The two-tailed *P* value is 0.0025.

6. Barwick, J. L. *et al.* Trans-species gene transfer for analysis of glucocorticoid-inducible transcriptional activation of transiently expressed human CYP3A4 and rabbit CYP3A6 in primary cultures of adult rat and rabbit hepatocytes. *Mol. Pharmacol.* **50**, 10–16 (1996).
7. Kliewer, S. A. *et al.* An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. *Cell* **92**, 73–82 (1998).
8. Blumberg, B. *et al.* SXR, a novel steroid and xenobiotic-sensing nuclear receptor. *Genes Dev.* **12**, 3195–3205 (1998).
9. Lehmann, J. M. *et al.* The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions. *J. Clin. Invest.* **102**, 1016–1023 (1998).
10. Bertilsson, G. *et al.* Identification of a human nuclear receptor defines a new signaling pathway for CYP3A induction. *Proc. Natl Acad. Sci. USA* **95**, 12208–12213 (1998).
11. Wrighton, S. A. *et al.* Demonstration in multiple species of inducible hepatic cytochromes P-450 and their mRNAs related to the glucocorticoid-inducible cytochrome P-450 of the rat. *Mol. Pharmacol.* **28**, 312–321 (1985).
12. Schuetz, E. G. & Guzelian, P. S. Induction of cytochrome P-450 by glucocorticoids in rat liver. II. Evidence that glucocorticoids regulate induction of cytochrome P-450 by a nonclassical receptor mechanism. *J. Biol. Chem.* **259**, 2007–2012 (1984).
13. Jones, S. A. *et al.* The pregnane X receptor: a promiscuous xenobiotic receptor that has diverged during evolution. *Mol. Endocrinol.* **14**, 27–39 (2000).
14. Savas, U., Griffin, K. J. & Johnson, E. F. Molecular mechanism of cytochrome P-450 induction by xenobiotics: an expanded role for nuclear hormone receptors. *Mol. Pharmacol.* **56**, 851–857 (1999).
15. Waxman, D. J. P450 gene induction by structurally diverse xenobiotics: central role of nuclear receptors CAR, PXR, and PPAR. *Arch. Biochem. Biophys.* **369**, 11–23 (1999).
16. Blumberg, B. & Evans, R. M. Orphan nuclear receptors—new ligands and new possibilities. *Genes Dev.* **12**, 3149–3155 (1998).
17. Pinkert, C. A., Ornitz, D. M., Brinster, R. L. & Palmiter, R. D. An albumin enhancer located 10 kb upstream functions along with its promoter to direct efficient, liver-specific expression in transgenic mice. *Genes Dev.* **1**, 268–276 (1987).
18. Schuetz, E. G., Schinkel, A. H., Relling, M. V. & Schuetz, J. D. P-glycoprotein: a major determinant of rifampicin-inducible expression of cytochrome P4503A in mice and human. *Proc. Natl Acad. Sci. USA* **93**, 4001–4005 (1996).
19. Kolars, J. C., Schmiedlin-Ren, P., Shuetz, J. D., Fang, C. & Watkins, P. B. Identification of rifampin-inducible P450III4 (CYP3A4) in human small bowel enterocytes. *J. Clin. Invest.* **90**, 1871–1878 (1992).
20. Selye, H. Hormones and resistance. *J. Pharm. Sci.* **60**, 1–28 (1971).
21. Kolars, J. C., Benedict, P., Schmiedlin-Ren, P. & Watkins, P. B. Aflatoxin B1-adduct formation in rat and human small bowel enterocytes. *Gastroenterology* **106**, 433–439 (1994).
22. Bocherding, S. M., Baciewicz, A. M. & Self, T. H. Update on rifampin drug interactions. II. *Arch. Intern. Med.* **152**, 711–716 (1992).
23. Hebert, M. F., Roberts, J. P., Prueksaritanont, T. & Benet, L. Z. Bioavailability of cyclosporine with concomitant rifampin administration is markedly less than predicted by hepatic enzyme induction. *Clin. Pharmacol. Ther.* **52**, 453–457 (1992).
24. Burger, H. J., Schuetz, J. D., Schuetz, E. G. & Guzelian, P. S. Paradoxical transcriptional activation of rat liver cytochrome P-450 3A1 by dexamethasone and the antiglycorticoid pregnenolone 16 carbonitrile: Analysis by transient transfection into primary monolayer cultures of adult rat hepatocytes. *Proc. Natl Acad. Sci. USA* **89**, 2145–2149 (1992).
25. Hashimoto, H. *et al.* Gene structure of CYP3A4, and adult-specific form of cytochrome P450 in human livers, and its transcriptional control. *Eur. J. Biochem.* **218**, 585–595 (1993).
26. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59 (1989).
27. Xie, W., Chow, L. T., Paterson, A. J., Chin, E. & Kudlow, J. E. Conditional expression of erbB2 oncogene in transgenic mice elicits striking hyperplasia in stratified epithelia, and up regulation of TGF α expression. *Oncogene* **18**, 3593–3607 (1999).
28. Yanagimoto, T., Itoh, S., Muller-Enoch, D. & Kamataki, T. Mouse liver cytochrome P-450 (P450III4A1): its cDNA cloning and inducibility by dexamethasone. *Biochim. Biophys. Acta* **1350**, 329–332 (1992).
29. Jelinek, D. F., Andersson, S., Slaughter, C. A. & Russell, D. W. Cloning and regulation of cholesterol 7 α -hydroxylase, the rate-limiting enzyme in bile acid biosynthesis. *J. Biol. Chem.* **265**, 8190–8197 (1990).
30. Kimura, S., Gonzalez, F. J. & Nebert, D. W. Mouse cytochrome P3-450: complete cDNA and amino acid sequence. *Nucleic Acids Res.* **12**, 2917–2928 (1984).

Acknowledgements

We thank K.-F. Lee for ES cells and his advice on ES cultures; Y. Dayn for her assistance in DNA and ES cell microinjections; A. Pierce, H. Juguilon, G. Nelson and I. Ong for technical assistance; J. Simon and M. Lieberman for assistance in photography; and E. Stevens and L. Ong for administrative assistance. W.X. is supported by Susan G. Komen Breast Cancer Foundation. M.D. is a C.J. Martin Research fellow of Australia. R.M.E. is an Investigator of the Howard Hughes Medical Institute at the Salk Institute for Biological Studies and March of Dimes Chair in Molecular and Developmental Biology. This work was supported by Mathers Foundation (R.M.E.), the Howard Hughes Medical Institute (R.M.E.) and grants from the NIH (P.S.G.).

Correspondence and requests for materials should be addressed to R.M.E. (e-mail: evans@salk.edu).

correction

The protein kinase Pak3 positively regulates Raf-1 activity through phosphorylation of serine 338

Alastair J. King, Huaiyu Sun, Bruce Diaz, Darlene Barnard, Wenyan Miao, Shubha Bagrodia & Mark S. Marshall

Nature **396**, 180–183 (1998)

In this letter, we stated that the isoform of p21-activated protein kinase (PAK) purified from rat spleen was Pak3. At the time of going to press, this was correct nomenclature for the rat isoform based on the SwissProt protein sequence database. However, under the restructuring of PAK nomenclature within this database (December 1998) the isoform we had previously purified has now been classified as Pak2. Although we were able to detect phosphorylation of the catalytic domain of Raf-1 (CR3) by the purified kinase, now identified as Pak2, we now note that experiments using recombinant protein (Fig. 3c, d) or DNA constructs (Fig. 4) all used *bona fide* Pak3 (murine) from a qualified source (R. Cerione laboratory). This suggests the potential involvement of various Pak isoforms in the regulation of Raf-1 activity through phosphorylation of Ser 338. □

erratum

Turbulent convection at very high Rayleigh numbers

J. J. Niemela, L. Skrbek, K. R. Sreenivasan & R. J. Donnelly

Nature **404**, 837–840 (2000)

In this Article the bold paragraph incorrectly stated that thermal transport had been investigated over the range $10^6 \leq Ra \leq 10^7$. It should have read 'Here we investigate thermal transport over eleven orders of magnitude of the Rayleigh number ($10^6 \leq Ra \leq 10^{17}$)'. □

Oligo, oligo, oligo

Oligonucleotides form the core of this week's selection.

Versi-Dry

Nalgene

www.nalgenenunc.com

Spill containment in the lab

Versi-Dry Lab Soakers, provided in dispenser packs, rolls and mats, contain thousands of 'Thirsty Cells' that quickly absorb spills and cushion breakable items. The soakers have a waterproof chemical-resistant polyethylene backing that also grips surfaces and equipment. They can also be used in a tray, drawer or as a shelf liner to help cushion glassware and other delicate items from breaking when dropped or tipped over on hard surfaces. Soft and non-abrasive, Versi-Dry Lab Soakers won't scratch delicate surfaces or equipment. Standard Absorbent Lab Soakers can absorb up to 750 ml of liquid per square metre. Super Absorbent Lab Soakers will absorb up to 1,050 ml per square metre.

Reader Service No. 100



Versi-Dry: making spilling things fun?

output directly into an Excel spreadsheet to allow calculation of the melting point.

Reader Service No. 102

CEQ 2000XL

Beckman Coulter

www.beckmancoulter.com

Faster runs, longer reads

The CEQ 2000XL DNA Analysis System provides automated DNA sequencing and fragment analysis, using capillary electrophoresis technology with a linear polyacrylamide gel to read out up to 700 bases, and a cycle time per capillary of 100 minutes. Plate washing, gel casting and sample loading are all eliminated with this system, which also features laser-induced fluorescence detection, four-colour dideoxycytosine chemistry and a microplate format. A software program for accuracy determination is included.

Reader Service No. 101

Melting point accessory

Hitachi Scientific Instruments

www.hsi-europe.com

Eliminates need for theoretical calculations

The melting point of an oligonucleotide is an important parameter in determining the suitability of the material for specific reactions, since double-strand forms can convert into the single-strand form. This melting point accessory for HSI UV/Vis spectrophotometers allows direct measurement using a thermocycler with Peltier cooling and PCR controller. A standard sample cuvette is fitted with a temperature probe. Software controls the heating and cooling cycle, providing temperature control of 0.5 °C per min over a range of 20–80 °C. Absorbance measurements are made at the optimum wavelength of 260 nm for DNA during the temperature cycle, with absorbance and temperature measurements

Versitag

NEN Life Science

www.nen.com

Chemical labelling systems

Two new labelling systems are available from NEN for non-enzymatic, non-radioactive labelling of DNA, RNA and oligonucleotides. The technology incorporates a platinum moiety that acts as a linker between the nucleic acid and the hapten. A single kit can be used to label RNA, DNA and oligonucleotides. Both systems can scale up or down the amount of probe being labelled — and neither uses hazardous chemicals.

Reader Service No. 103

UIBC kit

Wako

www.wakousa.com

Iron binding, spectrophotometrically

The L-Type assay for unsaturated iron binding capacity (UIBC) is based on bathophenanthroline (BPT) as a chromogen. When a sample is mixed with buffer containing a known excess amount of iron, unsaturated transferrin in serum quantitatively associates with iron in the buffer and is converted to a saturated state. The portion of iron that remains unassociated is assayed through colour development with ascorbic acid and the disulphonic acid disodium salt of BPT. The unsaturated transferrin concentration in the sample can be determined by calculating the decreased amount of iron in the buffer. UIBC concentrations are measured spectrophotometrically at 545 nm. The new reagent is liquid, ready-to-use and stable on storage. The assay is highly specific, with no significant interference from ascorbic acid, bilirubin, haemoglobin or intralipids.

Reader Service No. 104

Perfectprep Plasmid Kit

Eppendorf

www.eppendorf.com

Fewer steps to plasmid DNA

The Perfectprep Plasmid Kit from Eppendorf is designed to speed up minipreps, with even fewer handling steps than needed previously. This new version of the Mini Kit achieves the isolation of ultra-pure plasmid DNA within 20 minutes, producing a product suitable for the usual downstream procedures. Automatic sequencing produces extremely long read lengths with low error rates. High-reproducibility yields are obtained via improved alkaline lysis in combination with an effective binding



A good read: Miniprep for plasmid DNA production.

new on the market

matrix. The plasmid DNA can be used directly after elution from the column, with no further precipitation steps necessary. Yields of up to 20 µg plasmid DNA can be obtained. Handling the kit in spin-column format is simplified as no organic solvents are used. The Perfectprep Plasmid Kit is also available in a Midi, Maxi and XL format.

Reader Service No. 105

Polyplex

GeneMachines www.genemachines.com

Oligonucleotide synthesizer

This apparatus is able to synthesize 96 20-mers in under three-and-a-half hours and has a capacity of four plates or 384 oligos per day. Flushing of reagent lines is eliminated with parallel dispensing of reagents, reducing synthesis time and reagent consumption. The synthesis chamber provides an inert-gas environment, and synthesis progress can be monitored using full-plate tritium collection after any base addition.

Reader Service No. 106

ZipTip

Millipore www.millipore.com

Peptide elution, from femtomoles to picomoles

The ZipTip_{µ-C18} pipette tip for concentrating and desalting peptides and oligonucleotides contains a 0.2-µl bed of C₁₈ resin which allows elution of purified peptides and proteins in less than 1 µl. The pipette tip is placed on a single- or multi-channel pipettor, 22-gauge blunt-end HPLC needle, or automated sample preparation station. The sample is bound by aspirating and dispensing through the media. Suggested applications include mass spectrometry, microbore HPLC and capillary electrophoresis.

Reader Service No. 107

BioProbe

Enzo Diagnostics www.enzobio.com

In situ hybridization

This line of non-radioactive kits and reagents can be used for labelling DNA, RNA and oligodeoxynucleotide probes with biotin, fluorescein or other labels. They are applicable to *in situ* hybridization assays using cells and tissue samples.

Reader Service No. 108

DNAQuant

Promega www.promega.com

Devilishly clever

This system measures DNA using a series of enzymes that produce varying amounts of ATP, depending on the amount of DNA present. A luciferase/luciferin reaction is used to generate a light signal which reflects the amount of ATP produced. In the linear

range of the assay, the amount of light is proportional to the amount of DNA present. The sensitivity limit is 20 pg DNA.

Reader Service No. 109

DNA repair enzymes

Trevigen www.trevigen.com

Investigative tools for UV-damaged sequences

Two DNA repair enzymes specific for UV-damaged DNA sequences are now available. The T4 endonuclease V (denV) enzyme recognizes cyclobutane pyrimidine dimers, cleaves the N-glycosyl bond of the 5' thymine of the dimer and subsequently the phosphodiester backbone, leaving an unsaturated aldehyde. The chloroella virus PDG also recognizes cyclobutane pyrimidine dimers and excises the damaged bases, but, unlike the denV enzyme, cvPDG can cleave the *trans*-syn II photoisomer. The enzymes are purified and provided with an appropriate reaction buffer. The enzymes are suggested for use in investigations of UV irradiation damage to cells.

Reader Service No. 110

North2South

Pierce www.piercenet.com

Just add water

The manufacturers of this non-radioactive labelling and detection kit promise that users will be able to go from unlabelled probe to chemiluminescent detection in two hours with no loss of sensitivity. The one-step procedure can be performed in 15 minutes, with labelled probes lasting for a year. The kit contains all reagents required to label a probe with HRP and detect chemiluminescent substrate along with all necessary hybridization and wash buffers. The user need only supply the nucleic acid and some water.

Reader Service No. 111

Antibody service

Genovac www.genovac.de

Use the direct route from genes to antibodies

While most conventional immunization protocols necessitate protein purification or peptide synthesis, this new service from Genovac offers custom-made antibodies by genetic immunization with tested specificity directly with a cDNA clone. High-affinity antibodies can be generated against native proteins, which are synthesized within the immunized animal. Sequence analysis is carried out to determine the optimal immunization protocol and the cDNA recloned into a specially designed vector for immunization and testing. Rabbits and/or mice are immunized with the cDNA and the immune sera or hybridoma supernatants tested to identify specific antibodies. Hybridomas and purified monoclonal antibodies are supplied three to four months after sequence submission.

Reader Service No. 112

E-Gels

Invitrogen www.invitrogen.com

Make a new resolution

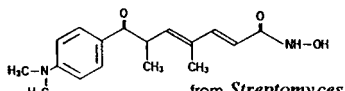
E-Gels are bufferless, pre-cast agarose gels that are intended to virtually eliminate set up because they contain everything you need for fast and convenient electrophoresis. With the efficient separation of large DNA in view, E-Gels are now available in the widely used 0.8% agarose concentration. The 0.8% E-Gel resolves DNA fragments across a wide range of sizes, from 800 bp to 10 kb.

Reader Service No. 113

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

Trichostatin A



from *Streptomyces*

Catalog No. 204-11991

This histone deacetylase inhibitor is highly effective at low concentrations, with high specificity. For more information, please call

Wako BioProducts: (800) 992-9256; (804) 271-7677.

Custom Peptides & Polyclonal Antibodies

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

Research Genetics
U.S. or Canada 800-533-4363
FAX 256-536-9016

Check out our homepage at
<http://www.resgen.com>

READER ENQUIRY NO. 98

READER ENQUIRY NO. 39